

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representation of  
The original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**

**PCT**WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification 5 :</b> <b>C12N 15/10, C12Q 1/68</b>	<b>A1</b>	<b>(11) International Publication Number:</b> <b>WO 91/05861</b> <b>(43) International Publication Date:</b> 2 May 1991 (02.05.91)
<b>(21) International Application Number:</b> PCT/US90/05938 <b>(22) International Filing Date:</b> 16 October 1990 (16.10.90)  <b>(30) Priority data:</b> 422,392 16 October 1989 (16.10.89) US  <b>(71) Applicant:</b> GENELABS INCORPORATED [US/US]; 505 Penobscot Drive, Redwood City, CA 94063 (US).  <b>(72) Inventors:</b> LOVETT, Michael ; 416 Durant Way, Mill Valley, CA 94941 (US). REYES, Gregory, R. ; 2112 St. Francis Drive, Palo Alto, CA 94303 (US). WEISSMAN, Sherman, M. ; 495 St. Ronan Street, New Haven, CT 06511 (US). JORGENSEN, Linda, M. ; 555 Addison Avenue, Palo Alto, CA 94301 (US).		<b>(74) Agent:</b> DEHLINGER, Peter, J.; 350 Cambridge Avenue, Suite 100, Palo Alto, CA 94306 (US).  <b>(81) Designated States:</b> AT (European patent), AU, BE (European patent), CA, CH (European patent), DE (European patent), DK (European patent), ES (European patent), FR (European patent), GB (European patent), GR (European patent), IT (European patent), JP, KR, LU (European patent), NL (European patent), SE (European patent).  <b>Published</b> <i>With international search report.</i> <i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>
<b>(54) Title:</b> NON-SPECIFIC DNA AMPLIFICATION  <b>(57) Abstract</b>  A method of non-specifically amplifying different-sequence fragments in a mixture of duplex DNA fragments is disclosed. The fragments are provided with end linkers, and the mixture is amplified by successive primer-initiated replication. Also disclosed is a method of cloning cDNA species which are homologous to a region of contiguous genomic DNA and are selected from a mixture of cDNA species.		

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	ES	Spain	MG	Madagascar
AU	Australia	FI	Finland	ML	Mali
BB	Barbados	FR	France	MR	Mauritania
BE	Belgium	GA	Gabon	MW	Malawi
BF	Burkina Faso	GB	United Kingdom	NL	Netherlands
BG	Bulgaria	GR	Greece	NO	Norway
BJ	Benin	HU	Hungary	PL	Poland
BR	Brazil	IT	Italy	RO	Romania
CA	Canada	JP	Japan	SD	Sudan
CF	Central African Republic	KP	Democratic People's Republic of Korea	SE	Sweden
CG	Congo	KR	Republic of Korea	SN	Senegal
CH	Switzerland	LI	Liechtenstein	SU	Soviet Union
CI	Côte d'Ivoire	LK	Sri Lanka	TD	Chad
CM	Cameroon	LU	Luxembourg	TG	Togo
DE	Germany	MC	Monaco	US	United States of America
DK	Denmark				

NON-SPECIFIC DNA AMPLIFICATION

The present invention is a continuation-in-part of U.S.  
5 patent application for "DNA Amplification and Subtraction  
Techniques," Serial No. 208,512, filed June 17, 1988.

**Field of the Invention**

The present invention relates to a method of non-specific  
10 DNA amplification, and to applications of the method for cDNA  
selection.

**References**

- Aegerter-Shaw, M.J., Immunol 38, 3688 (1987)  
15 Azen, E.A., et al., Science 226, p. 967 (1984)  
Azume, C., et al., Nucl. Acid Res. 14, 9149-9158 (1986)  
Barsh, S., et al., Nature 310, p. 650 (1984)  
Campbell, H. D. et al., Proc. Natl. Acad. Sci. USA 84,  
p. 6629 (1987)  
20 Carle, G. F. and M. V. Olson, Proc. Natl. Acad. Sci. USA  
82, 3756-3760 (1985).  
Carle, G. F. and M. V. Olson, Nucleic Acids Res. 12,  
5647-5664, (1984).  
Collins, F. S. et al., Science 235, p. 1046 (1987)  
25 Chomzynski, P. et al., Anal. Biochem. 162, 156-159  
(1987).  
Chu, G. et al., Science 234, p. 1582 (1986)  
Church, G.M., and Gilbert, W., Proc Nat Acad Sci USA 81  
1991-1995 (1984)  
30 Clark, S.C., and Kamen, R., Science 234, p. 129 (1987)  
D'Eustachio P., et al., J. Immunol. 137, p. 3990 (1986)  
Erickson, B.W., et al., Proc Nat Acad Sci USA 81, p.  
7171 (1984)

- Hames, B.D. and Higgins, S.J. eds., "Nucleic Acid Hybridization: A Practical Approach," IRL Press, Oxford (1985)
- Huebner, K., et al., Science 230, p. 1282 (1985)
- Kaledin, A.S., et al., Biokhimiya 45, 644-651 (1980)
- 5 Karin, M., et al., Proc Natl Acad Sci USA 81, p. 5494 (1984)
- Kashmiri, S. V. S. et al., J. Virology 49, 583-589 (1984)
- Kiyokawa, T. et al., Proc. Natl. Acad. Sci. USA 81,  
10 6202-6206 (1984)
- Kohne, D.E., et al., Biochemistry 16(2), p. 5329 (1977)
- LeBeau, M.M., et al., Proc. Natl Acad Sci USA 84, p. 5913 (1987)
- LeBeau, M.M., et al., Blood 73, 647-650 (1989)
- 15 Leonard, W.J., et al., Science 228, 1547 (1985)
- Maniatis, T. et al., Molecular Cloning, Cold Spring Harbor Press, (1982).
- Michiel, F., et al., Science 236, 1305-1308 (1987)
- Old, R.W., Woodland, H.R., Cell 38, p. 624 (1984)
- 20 Overhauser, J., et al., Nucl. Acid Res. 15 4617-4627 (1987)
- Poustka, A. et al., Genomics 2, 337-345 (1988)
- Ricciard, R. P. et al., Proc. Natl. Acad. Sci. USA 76, 4927-4931 (1979).
- 25 Sagata, N. et al., Proc. Natl. Acad. Sci. USA 82, 677-681 (1985).
- Scharf, S.J., et al., Science 233, p. 1076 (1986)
- Schrader, J. W. et al., Proc. Natl. Acad. Sci. USA 83,  
p. 2458 (1986)
- 30 Sealey, P.G., et al., Nucleic Acids Res. 13, p. 1905 (1985)
- Sherrington, R. et al., Nature 336, p. 164 (1988)

- Smith, C. et al., Methods in Enzymology 151, p. 461 (1987)
- Southern, E., Methods in Enzymology 68, 152-176, (1979).
- Thompson, J., et al., Anal Biochem 303, p. 334 (1987)
- 5     Tonegawa, S., Nature 302, p. 575 (1983)
- VanLeeuwen, B.H., et al., Blood 73, 1142-1148 (1989)
- Wasmuth, J. J. et al., Am J Human Genet 39, p. 397 (1986)
- Westin, G., et al., Proc Nat Acad Sci USAZ 81, p. 3811
- 10    (1984)
- Weiss, E.H., et al., Nature 310, p. 650 (1984)
- Wong, G.G., et al., Science 228, p. 810 (1985)
- Woo, S. L. C., Methods in Enzymology 68, 389-395 (1979)
- Yang, Y. C. et al., Blood 71 p. 958, (1988)
- 15     Young, R. A. and R. W. Davis, Proc. Natl. Acad. Sci. USA 80, 1194-1198 (1983).

#### Background of the Invention

The ability to isolate and clone cDNA fragments which are

20 homologous to contiguous genomic regions has important applications in medicine. As one example, isolating cDNAs which are all homologous to a DNA section containing a known but as yet unlocalized disease-causing gene would result in the isolation of that gene and flanking genes; this analysis would

25 facilitate the process of localization and analysis of the disease-causing gene. A second important application is the possibility of identifying new genes which are linked to known markers.

As another example, homologies between two related virus-

30 es may be utilized by binding to the genome of a first virus, cDNAs which are derived from a cell type infected with a second virus, then isolating those cDNA sequences which are homologous to the first viral genome.

Another hybridization method which has important applications to medical research involves the selection of genomic or cDNA fragments which are either common or unique to two different biological sources, e.g., cell types, or cell types at  
5 different stages of activation or development.

Standard methods for obtaining cDNAs on the basis of their hybridization to genomic DNA material are often limited in the amount of genomic DNA or cDNA material which is available, particularly when it is desired to clone a large number  
10 of different-sequence fragments obtained from the hybridization step. This limitation can result in loss of a large number of fragments whose abundance in the hybridization elution mixture is very low.

#### 15 Summary of the Invention

It is therefore one object of the invention to provide a method of cloning cDNA species which are homologous to a region of contiguous genomic DNA and are selected from a mixture of cDNA species. The method involves obtaining the  
20 region of interest of contiguous genomic DNA and preparing cDNA species which contain end-terminal priming sequences. Single-stranded cDNA species are isolated on the basis of their hybridization to the region of contiguous genomic DNA. These isolated single-stranded cDNA species are mixed with DNA  
25 polymerase, all four deoxyribonucleotide triphosphates, and primers homologous to the cDNA end-terminal priming sequences. This mixture is then reacted under conditions to produce sequence-independent amplification of the single-stranded cDNA species. The amplified cDNA species resulting from this  
30 reaction are cloned into a vector.

The method of the present invention can also include a second cDNA molecule isolation step. The single-stranded cDNA molecules are isolated on the basis of their homology to a

first set of selected genomic fragments and are then amplified. The amplified cDNA species are denatured and a new set of single-stranded cDNA species are isolated on the basis of their hybridization to a second set of selected genomic fragments. For example, the cDNA species may be rehybridized against the first set of selected genomic fragments for enrichment, or, hybridized against a different set of genomic fragments to select a subset of the cDNA species.

In the preparation of cDNA species to contain end-terminal priming sequences, the above method may further include the amplification of the cDNA species by sequence-independent amplification.

The region of contiguous DNA in the above method may contain a known gene or region of interest. DNA fragments containing the known gene or region of interest may be obtained as follows. A DNA section containing the region of interest is size fractionated, the fraction containing the known coding sequence is identified, and the DNA from the identified fraction is isolated. One method to identify the DNA fragments of interest is by primer-initiated amplification in the presence of primers which are homologous to sequences in the region of interest. This isolated DNA can then be digested with a restriction endonuclease to cut the region of interest into a plurality of smaller fragments.

In one embodiment of the above method, the contiguous genomic DNA section is bound to a solid support.

It will be appreciated that the contiguous genomic DNA section can be contained in a number of cloning vectors, known to one of ordinary skill in the art, for cloning large segments of DNA, for example, yeast artificial chromosomes.

One embodiment of the above method utilizes a contiguous genomic DNA section from the bovine leukosis virus genome and



a cDNA library made from mRNA obtained from a cell line infected with a virus derived from cell line 10C9.

Another application of the above method includes the means for identifying cDNAs which correspond to genes located in the same chromosomal region as a known gene or region of interest. The fragments containing the genomic DNA region containing a known gene or region can be obtained by preparative size fractionation of genomic DNA fragments containing the known gene. One method to identify the DNA fragments of interest is by adding primers homologous to the known gene, DNA polymerase, and all four deoxyribonucleotides, to the gel matrix and treating the gel matrix under conditions which promote amplification of the region of the known gene defined by the primers. Linkers, which are useful as primers for sequence-independent amplification, are then ligated to the ends. Redundant linkers on the ends of the molecules can be removed by digesting with the appropriate restriction enzyme. The DNA fragments are mixed with DNA polymerase, all four deoxyribonucleotides, and primers homologous to the linkers present on the ends of the DNA fragments, and reacted under conditions to produce sequence-independent amplification of the DNA fragments.

The above described method can be used for genomic DNA regions containing such known genes as growth factor genes and growth factor receptor genes including the following: interleukins (e.g., IL-5 and IL-3), GM-CSF, and erythropoietin. With each of these genes an appropriate cDNA library is chosen, such as, a T-cell cDNA library when using IL-5, or a renal cDNA library when using erythropoietin.

These and other objects and features of the invention will be more fully understood when the following detailed description of the invention is read in conjunction with the accompanying drawings.

**Brief Description of the Drawings**

Figure 1 is a flow diagram of the duplex amplification method of the invention;

Figures 2A-2C show the sequences of three exemplary  
5 linkers used in practicing the invention, after attachment to opposite ends of blunt-ended duplex fragments (solid lines);

Figure 3 illustrates one method for biotinylating amplified cDNA fragments, in practicing the invention;

Figure 4 is a flow diagram of the method of the invention  
10 for isolating DNA fragments which are unique to one fragment source;

Figure 5 illustrates the isolation method shown in Figure 4 applied to isolation of an RNA-derived sequence which is unique to individuals infected with a given viral agent;

15 Figure 6 illustrates the amplification method of Figure 1, as it is used for confirming the presence of ET-NANB viral agent in bile from an infected animal, and for assaying the presence of the viral agent in infected individuals;

Figure 7 shows regional assignments of potential growth  
20 factor/receptor genes within a subregion of chromosome 5. The following designations are used:

	CSF-GM,	granulocyte-macrophage colony stimulating factor;
	CD14,	encodes an antigen which is a surface marker of monocytes;
25	IL-3,	multi-colony stimulating factor;
	IL4,	B-cell growth factor;
	IL5,	T-cell replacing factor;
	PDGFR,	platelet-derived growth factor receptor;
30	ADRBR,	beta-2 adrenergic receptor;
	FGFA,	fibroblast growth factor;
	CSF1R,	CSF1 receptor (the protooncogene c- <u>fms</u> );
	CSF1,	macrophage colony stimulating factor;

Figure 8 shows the steps in the preparation of small amplified genomic fragments derived from a region of human chromosome 5 containing the human IL-5 gene;

Figures 9A and 9B show isolated genomic DNA from the HHW105 hamster/human hybrid cell line digested with NotI fractionated on a contour-clamped homogeneous electric field (CHEF) gel (9B), and sample products amplified with IL-5 specific primers (9B);

Figure 10 shows Southern blots of fragments amplified by IL-5-specific primers, which were fractionated on an agarose gel;

Figure 11 is an agarose gel fractionation of the non-specific amplification products of genomic fragments derived from the chromosome 5 region surrounding the IL-5 coding region;

Figure 12 is an agarose gel fractionation of T-cell cDNA's which hybridized with the genomic DNA fragments from Figure 11;

Figure 13 shows the hybridization of T-cell cDNA's isolated on the basis of homology to the IL-5 region of human chromosome 5 to a variety of probes, including probes specific for actin, IL-3 and IL-5 coding regions;

Figure 14 shows the hybridization of T-cell cDNAs isolated on the basis of homology to the IL-3 region of human chromosome 5 to a variety of probes including probes specific for IL-3, GM-CSF and actin coding regions, and to probes for total human genomic DNA;

Figure 15 shows long-range restriction maps, adapted from VanLeeuwen et al (1989), of the human chromosome 5 regions containing IL-5 and IL-3; and

Figure 16 shows how the method of the invention may be applied to identifying specific disease-related loci and their gene products.

## Detailed Description of the Invention

### I. Amplifying DNA Fragments

#### A. Sources of DNA Fragments

5        Figure 1 illustrates the method of amplifying duplex DNA fragments according to the invention. The duplex fragments are typically present in a fragment mixture, and typically a mixture of cDNA fragments produced from messenger RNA (mRNA) transcript species, although genomic DNA fragments or line-  
10    rized vectors or vector fragments are also suitable.

      Methods for isolating mRNA species from cellular or body-fluid samples, such as serum or bile, are well known. One method involves formation of a vanadyl-RNA complex, extraction of protein with chloroform/phenol, and precipitation  
15    with cold ethanol. In a second method, the RNA is extracted from a guanidinium isothiocyanate mixture with phenol, followed by a chloroform:isoamyl alcohol extraction, and precipitation of RNA from the aqueous phase with cold ethanol. The reader is referred to Maniatis, pp. 188-198, and references  
20    cited therein for details.

      As a rule, eukaryotic mRNAs are characterized by a 3' poly A terminal sequence which allows isolation by affinity chromatography, using oligo-dT or poly U bound to a solid support. In addition, or alternatively, total isolated RNA  
25    can be further fractionated by density gradient centrifugation, or agarose gel electrophoresis, to obtain a desired size fraction of RNA species.

      Production of duplex cDNAs from the isolated mRNA transcripts is by conventional oligo dT or random priming of first  
30    strand synthesis. The former method is advantageous where duplex cDNAs derived only from poly A RNAs are desired. In this method, the RNA transcripts are primed with oligo dT, to promote first-strand cDNA synthesis by reverse transcriptase

in the presence of all four deoxynucleotides. The 3' hairpin formed from the first-strand synthesis product is used to prime the synthesis of the second strand by E. coli DNA polymerase I, according to well-known methods (Maniatis, pp. 5 213216).

The random priming method of cDNA duplex formation is preferred where (a) duplex formation is not to be limited to poly A species, or (b) full-length duplex cDNAs are not required. The method for first-strand cDNA synthesis utilizes 10 an arbitrary sequence primer which is commercially available.

Following second-strand synthesis, the cDNA fragments are preferably blunt-ended, for example, by filling in sticky ends with the large fragment of E. coli DNA polymerase I (Klenow fragment) as described below in the Materials section.

15 Intact chromosomal DNAs can be isolated from agarose-immobilized cell preparations in those cases where chromosomal length DNAs are required. Genomic DNA from a selected cell source can also be isolated by standard procedures, which typically include successive phenol and phenol/chloroform 20 extractions with ethanol precipitation. Genomic DNA fragments, either in a fragment mixture or a purified preparation, are also suitable for amplification, according to the method of the invention. The duplex DNA is fragmented preferably by 25 striction endonucleases, although mechanical shearing may be employed. The fragmented genomic pieces may be size fractionated, or further treated to remove repetitive DNA. Other sources of double-stranded DNA fragments can include extra-chromosomal material, e.g., mitochondrial DNA, double-stranded 30 DNA viruses, or viruses which have as part of their life cycle a double-stranded intermediate, e.g., a retrovirus.

Cellular sources of genomic DNA fragments or RNA transcripts used for producing cDNA fragments include cultured

cell lines, or isolated cells or cell types obtained from tissue (or whole organs or entire organisms). Cell sources are of interest in a variety of subtraction techniques where it is desired to identify or isolate particular RNA transcripts or genomic material which are unique to one of two related cell sources. Body-fluid sources of DNA and/or mRNA transcripts are of interest primarily where the fluid is known or suspected to contain a viral agent or other microbe of interest. Example 4, for instance, describes cDNA fragment mixtures produced from RNA isolated from bile, taken before and after infection of cynomolgus monkeys with enterically transmitted non-A/non-B (ET-NANB) hepatitis virus.

Linearized or fragmented plasmid DNA, or fragmented phage DNA is another source of DNA fragments which one might wish to amplify. The vector DNA is obtained from purified plasmid or phage DNA according to conventional techniques, and linearized and/or fragmented by digestion with selected restriction endonuclease(s). Example 2 describes DNA fragments obtained by HaeIII digestion of phiX174 phage and linearized pIAN13 plasmid.

#### B. Fragment Linkers

According to an important feature of the invention, the duplex DNA fragments to be amplified are ligated at their opposite end to a linker, to provide a priming sequence for strand duplication, i.e. an end-terminal primer. The linker is preferably a short duplex DNA fragment, typically about 20-30 basepairs in length, having a defined base pair sequence. One end of the linker is designed for enzymatic ligation to opposite fragment ends. Where the fragments are blunt-ended cDNAs or genomic fragments, this linker end is also blunt, for ligation to the fragment ends in the presence of T4 DNA ligase, as described in Example 2. Where the DNA

fragment mixture is formed by endonuclease digestion of genomic DNA, and produces fragments with sticky ends, these ends can either be filled in to generate blunt ends or the linkers can be prepared with complementary sticky ends at the  
5 fragment-ligation end.

Since the ligation reaction involving the linkers also produces linker-to-linker ligation, the linkers are preferably designed so that dimers can be selectively cleaved by restriction endonuclease digestion. This can be done by constructing  
10 the blunt end of the linker with a sequence representing one-half of a selected restriction-site sequence. The restriction site formed by blunt-end dimerization is preferably a rare cutter site, so that digestion of the fragments with the associated endonuclease, after linker addition to the DNA  
15 fragments, does not cleave a significant number of fragments internally, and thereby produce fragments with a linker at one end only.

Figure 2 shows DNA fragments (solid lines) with exemplary linkers A, B, or C attached to opposite fragment ends.  
20 As seen, each of the three linkers has one-half of an NruI site at its blunt end, whereby linker dimers formed by blunt end ligation can be cleaved by NruI digestion. It should be noted that linker ligation to the DNA fragments is directional, and linker self-ligation is limited to dimer formation,  
25 due to the one staggered end of the linker. The linkers are joined to blunt-ended fragments as indicated above, and as illustrated at the top in Figure 1. It will be appreciated that analogous linkers with one sticky end for ligation to corresponding fragment sticky ends could be similarly employed.  
30 However, these linkers have the disadvantage that recleavage for removal of linker dimers is not possible.

Also as seen in Figure 2, the linkers preferably include one or more internal restriction sites at which the fragments

can be cleaved after amplification. The internal linker restriction site(s) can serve three purposes. First, where the amplified fragments are used in the DNA subtraction method described in Section D-1 below, the restriction site can be  
5 used to cleave away a major portion of the linker in one of the two fragment mixtures, to prevent hybridization between linker regions in strands from the two different mixtures, where the same linker is used in the amplification of both fragment mixtures. Secondly, the linker cutting site(s)  
10 allows the amplified fragments (or the hybridized fragments in the Section D1 subtraction method), to be equipped with desired sticky end sites for cloning into a cloning vector. Thirdly, as will be seen in Section D-1, the linker restriction site may be used to create a single-strand end, for  
15 purposes of biotinylating the fragments. In particular, an internal EcoRI site produces an AATT overhang which can be filled in with commercially available biotinylated dUTP and dATP nucleotides.

Synthetic duplex oligonucleotide linkers having selected  
20 sequences, such as those shown for the three exemplary linkers in Figure 2, can be prepared using commercially available automated oligonucleotide synthesizers. Alternatively, custom designed synthetic oligonucleotides may be purchased, for example, from Synthetic Genetics (San Diego, CA).

25

### C. Fragment Amplification

The method steps described above are shown at the top in Figure 1. These steps include obtaining duplex fragments, blunt-ending and attaching a linker to opposite fragments ends  
30 to produce linker-carrying fragments, such as illustrated in Figure 2, and treating with a selected endonuclease to cut linker multimers at their blunt-end junctions. Alternatively, the linker dimers can be removed by fractionating the DNA by



gel filtration or agarose gel electrophoresis, and eluting all DNAs larger than linker dimers.

The linker-fragments from above are amplified by the repeated fragment duplication according to the following steps. First, the fragments are mixed with a large molar excess of a single-strand oligonucleotide primer, typically a  $10^6$ - $10^9$  molar excess. The primer sequence is homologous to the fragment end linker; that is, the primer sequence and length is such as to promote hybridization to the complementary-linker regions of the two strands under moderately stringent reannealing conditions. Another requirement is that the primer, when hybridized to the linker region of the denatured fragment strands, be capable of priming polymerase-catalyzed strand replication; that is, the internal end of the primer provides a free 3'-OH. In the case of the linkers shown in Figure 2, preferred primer sequences are:

d(5'-GGAATTCGCGGCCGCTCG-3') for linker A;  
d(5'-GGTTGCGGCCGCTCG-3') for linker B; and  
20 d(5'-GACGCGTGAATTCTCG-3") for linker C.

Single-strand oligonucleotide primers having the desired sequence are prepared by conventional methods or can be obtained from commercial laboratories, as above.

25 The denatured fragments and primer are cooled, typically between about 37-60°C, to allow primer hybridization with the fragment linker ends. The cooling period is kept quite short, typically less than about 5 minutes, to minimize strand-strand hybridization which would be expected on long-term reannealing. At this point, or at the primer-addition stage, the four deoxynucleotides and a DNA polymerase capable of catalyzing  
30 second-strand, primed replication are added, and the reaction

mixture is brought to a temperature suitable for enzymatic strand replication.

In the method described in Example 2 below, the deoxynucleotides and DNA polymerase were added prior to fragment denaturation. After heat denaturing, the mixture was annealed at 50°C for two minutes, then brought to 72°C for 5-12 minutes for primed, second-strand replication. The DNA polymerase used is Thermus aquaticus DNA polymerase (Taq DNA polymerase), which is relatively heat-stable at up to 95°C for brief periods (Kaledin et al; Perkin Elmer Cetus).

It will be appreciated from the above, and from Figure 1, that the single added primer hybridizes with the 3'-end linker region in each DNA strand, and thus a doubling of fragment number occurs. The above replication procedure, which involves fragment denaturation by heating, cooling to form a fragment strand-primer complex, and second strand replication of the complex in the presence of DNA polymerase, is repeated until a desired concentration of fragments is achieved. In the above example, which employs a heat-stable Taq polymerase, the three replication steps are carried out simply by heating the fragment mixture to a denaturing temperature (above the  $T_m$  of the fragments), cooling briefly to allow fragment/ primer complex formation, and incubating for a period sufficient for second-strand synthesis.

Example 2 illustrates the application of the fragment amplification method to a linearized DNA plasmid (pIAN13) and to a fragment mixture produced by HaeIII digestion of phage phiX174 fragments.

#### 30 D. Utility

The amplification method is useful for DNA duplex fragment amplification where limited amounts of genomic or cDNA fragment material is available and/or as a simple method of

amplifying duplex DNA material where DNA sequences are unknown. For example, in the probe binding study reported in Example 4, the method was used to increase by several orders of magnitude, the number of cDNA fragments derived from bile infected with enterically-transmitted non-A/non-B hepatitis (ET-NANB) viral agent. The amplification allowed the relatively small amount of viral agent present in the bile to be easily identified as an amplified cDNA fraction.

The method is useful for amplifying the amount of a fragment or fragments available for cloning, to enhance cloning efficiency. One advantage of the method for cloning is the ability to provide the amplified fragments with end cloning sites suitable for the cloning vector which has been selected.

The method can be applied to purified, enriched, or subfractionated DNA fragments, for increasing the amount of fragment material available for further processing, e.g., further enrichment. For example, in the fragment-isolation method described in below, the isolated unique sequence fragments can be further enriched by additional amplification and selection. As another example, a DNA subfraction obtained by gel electrophoresis can be selectively amplified. Here the fragment mixture, prior to fractionation, is equipped with an end linker, and the selected fragment band is further isolated, either in situ within the gel, or after selective elution, by addition of primer, polymerase, and nucleotides, with repeated heating and cooling replication steps. The following specific applications are illustrative.

### 1. Isolating Unique Sequences

#### (a) Hybridizing Fragment Mixtures

Figure 4 is a flow diagram of the method for isolating unique sequences from one of two fragment mixtures, according

to the invention. The two fragment mixtures are referred to herein as positive- and negative-source mixtures, and are derived from positive and negative sources of nucleic acid species, respectively. The positive source may be any cell,  
5 cell group, tissue, organ, subcellular fractions, or other defined source of RNA transcripts or genomic fragments, including vector fragments, which contain a number of different-sequence species. The negative source is similar to the positive source, but lacks one or more positive source species.

10 By way of example, the positive and negative sources may be cell or tissue samples which are infected (positive source) and non-infected (negative source) with a microbial agent. Here the nucleic acid species associated with the infecting agent, either in the form of genomic material or RNA tran-  
15 script species, are unique to the infected (positive) source. Similarly, the positive and negative sources may be a body fluid, such as serum or bile, from infected and non-infected subjects, respectively. Where the present method is applied to the study of hereditary diseases, the positive source is  
20 typically a normal cell line or tissue capable of producing the full range of normal proteins associated with that cell, and the negative source is a genetically altered cell which is suspected of an altered or lost mRNA species. That is, the positive source will yield a normal mRNA species which is not  
25 present in the genetically altered, negative source.

As another example, it is known that malignant transformation in cells may be triggered by the activation of oncogenes, as evidenced by expression of new RNA transcripts associated with the malignant state. In applying the method  
30 of the invention to isolating and identifying such RNA species, the malignant cells or tissue serves as the positive source of unique (tumor-related) RNA species, and normal, non-transformed cells or tissue, as the negative source.

Other examples of negative and positive sources of transcripts or DNA duplex fragments are contemplated.

The positive and negative sources shown in Figure 4 yield mixtures of mRNA species, and these are converted to  
5 corresponding duplex DNA fragments as discussed in Section A above. The duplex fragments in the two mixtures are blunt-ended and separately ligated at opposite fragment ends to a linker, as described in Section B. The linker attached to the positive-source fragment is preferably different than that  
10 attached to the negative-source fragments. For example, in the method outlined in Example 2, the positive source fragments are ligated to linker A in Figure 2, and the negative-source material, with linker C. Attaching linkers with different sequences to the two fragment mixtures avoids the  
15 problem of linker hybridization when the two mixtures are hybridized. Alternatively, the same linker may be attached to both fragment mixtures, if the linker contains a restriction site which allows substantially all of the linker to be removed by restriction endonuclease digestion from the negative-source fragment mixture before hybridization.  
20

Each fragment mixture is then amplified by repeated replications, as described in Section 1C above. In a preferred embodiment, the amplified, negative-source fragment strands are biotinylated or otherwise equipped with a ligand  
25 which can be used to selectively bind and remove strands or duplex species containing the ligand, by affinity chromatography. A variety of methods for biotinylating duplex fragments are available. For example, biotin-labeled deoxynucleotides, such as Bio=11-UTP and Bio-7-dATP, can be included in  
30 the amplification reaction mixture during the final rounds of replication. Duplex fragments are readily photobiotinylated by a standard procedure, as outlined in the protocol supplied by the manufacturer, Clontech (Palo Alto, CA). In one pre-

ferred method, the amplified fragments are treated with an endonuclease, such as EcoRI, which produces a 5'-AATT-3' sticky end. The sticky end is filled in with Klenow fragment in the presence of Bio-11-UTP and/or Bio-7-dATP, both of which  
5 are commercially available.

The two fragment mixtures are now combined and heat-denatured to single strand form. Preferably the negative-source fragments are present in substantial molar excess, typically about a tenfold excess, to insure that non-unique  
10 fragments in the positive-source mixture have a high probability of hybridizing with complementary strands from the negative-source fragments.

The positive- and negative-source fragments are hybridized under conditions in which hybridization between the  
15 linkers in the two different mixtures does not occur. This may be accomplished, as noted above, by using different sequence linkers in the two mixtures, by cleaving the linkers from the negative- source fragment mixture prior to hybridization, or by hybridizing the fragment strands under conditions  
20 which preclude linker hybridization. Alternatively, the hybridization mixture may contain a large molar excess of primer oligonucleotides effective to hybridize with the positive-source linker end regions, thus preventing inter-strand linker hybridization.

25 The hybridization reaction may be carried out according to classical hybridization techniques, such as detailed in Hames, or more preferably, by rapid hybridization techniques, such as the phenol emulsion reassociation technique (PERT), as described by Kohne, or the guanidinium thiocyanate method, as  
30 described by Thompson.

In the PERT hybridization method, the fragments are hybridized at a final DNA concentration of preferably between about 5-50 ug/ml in a phenol reaction mixture, and the reac-

tion is carried out at 60-70°C for up to 5-6 hours. Alternatively, the reaction can be carried out at room temperature in the presence of formamide. The rate of annealing can be followed by hyperchromic shift at 280 nm or by diluting an aliquot of the reaction mixture with buffer, and quickly passing the material through a hydroxyapatite (HAP) column (Kohne).

In the guanidine thiocyanate method, the fragments are hybridized at a final DNA concentration of preferably between about 5-50 ug/ml in a 4M guanidine thiocyanate solution containing 8 mM DTT, and 20 mM Na citrate, pH 5.8. The mixture is heated to 65°C for 5 minutes, to insure complete denaturation, then incubated at 26°C for periods up to 72 hours, until complete hybridization has occurred (Thompson). The formation of DNA/DNA hybrids can be confirmed as above.

(b) Isolating Unique-Sequence Fragments

Following hybridization, the reaction components are contacted with a solid support effective to preferentially bind non-unique fragments. In a preferred embodiment, where the negative-source strands are biotinylated, the reaction components are separated on an avidin or streptavidin column which is effective to bind all biotinylated strands and non-biotinylated (positive-source) strands which are hybridized therewith. Since the biotinylated strands are present in large molar excess, virtually all common-sequence strands from the positive-source will be bound to a complementary, biotinylated strand, and thus removed by the affinity chromatography.

Methods for separating the hybridization mixture on a streptavidin column are detailed in Example 3. Briefly, the hybridization mixture is passed through a column packed with streptavidin-agarose and washed extensively with elution buffer. The eluates from several washes are pooled and the nucleic acid is concentrated by ethanol precipitation, yield-

ing a fragment mixture which is enriched for sequences which are unique to the positive-source material.

According to an important feature of the invention, the fragments isolated will contain the end linkers originally attached to the positive-source fragments, and thus can be further amplified, by the amplification method described in Section C. Where the positive-source linker is different than that on the negative-source fragments, additional amplification (employing the "positive-source" primer) will selectively amplify positive-source fragments, providing further enrichment of sequences which are unique to the positive source. As indicated in Figure 4, the isolated amplified material may be rehybridized with the negative source fragments (which may be additionally amplified if necessary), and unique fragments again isolated by affinity chromatography, to enhance enrichment.

The final isolated fragments can be treated with DNA polymerase to ensure complete strand replication in the duplex species, then treated with a selected linker-site endonuclease, to provide fragment ends suitable for cloning or the like. Alternatively, or in addition, the isolated fragments can be radiolabeled according to known procedures, for use as a DNA probe.

## 25 (2) Isolation and Cloning of Viral Agent

Figure 5 illustrates one application of the subtraction method, for isolation and cloning of a viral agent present as RNA transcript in an infected source, such as selected tissue or body fluid (e.g. bile). Here RNA mixtures isolated from infected (positive) and non-infected (negative) sources are used to produce cDNA duplex fragment mixtures, which are blunt-ended, ligated to a selected linker, and amplified as above. After biotinylating the negative-source fragments, the



two fragment mixtures are combined, denatured, reannealed, and separated by affinity chromatography (Section D-1). The isolated fragments, which contain a selected restriction site in their end-linker regions, are cut at this site and cloned  
5 in a suitable expression vector, such as the lambda gt11 vector. The resulting lambda gt11 phage constitute a fragment library enriched for sequences unique to the positive source material.

To select for desired isolates which produce a viral  
10 antigen, host cells are infected with library phage, and reacted with antiserum from an infected individual, to bind virus-related antibodies to host cells producing viral antigens. The cells are then washed to remove unbound antibody and contacted with a reporter-labeled anti-human IgG antibody,  
15 to label cells which have bound virus-specific antibodies. These cells are identified by the presence of the reporter. The vector(s) containing a viral insert can be used for continued antigen production, or as a source viral sequence probe.

20 Alternatively, the cloned isolated fragments in the fragment library can be identified and selected using DNA probes. By way of illustration, the fragment library generated as above can be replica plated and hybridized with labeled cDNA probes from infected and non-infected sources. Probe  
25 hybridization can be carried out by conventional Southern blotting. Those library vectors which contain a viral-agent insert will hybridize with viral-agent cDNA probes present in the positive-source cDNA probes, but not with negative-source probes.

### 3. Identification and Isolation of Cellular Nucleic Acid Products

Methods similar to the DNA subtraction and probe selection method just described can be applied to identifying and isolating nucleic acid fragments which are unique to cells expressing oncogenes, or cells deficient in transcripts related to genetic disease.

Figure 6 illustrates how the amplification technique described in Section C can be used to confirm that the unique-sequence fragment clone(s) identified as above are in fact unique to the positive source. Here the two cDNA fragment mixtures prepared from positive and negative sources are amplified, as in Section C, and fractionated by gel electrophoresis. The size fractionated gel fragments are then transferred by conventional Southern blotting to filters and hybridized with the isolated cloned insert of interest, after probe labeling. If the probe sequence is unique to the positive source fragments, hybridization will be seen with amplified positive-source fragments only; as indicated in Figure 6, the region of probe binding is fairly broad, as would be expected since the amplified fragments will have a range of sizes. This method of screening amplified cDNA fragments is detailed in Example 4, where the method has been used to confirm that an isolated ET-NANB fragment insert was unique to cDNA from an infected bile source.

### 4. Identifying cDNAs representing linked genes

There is no general rule for the chromosomal distribution of genes of like structure or function. However, gene duplication followed by divergence is a common mechanism for increasing the complexity and functional specialization of gene families. Although gene duplication could occur by transposition or retroposition, observations in a large number

of gene families suggest that duplications arose by tandemization of genes, perhaps as a result of unequal crossing over. Tandemization would create closely linked clusters of genes, which could then become separated by events such as chromosome  
5 breakage. However, even in cases where related gene are located on more than one chromosome, as in the genes of the complement system (Leonard et al.) or the genes for serum lipoproteins (Aegerter-Shaw), subsets of the gene system may occur as closely linked genes.

10 There are a large number of cases in which related genes are linked in clusters containing few unrelated intervening genes. Examples include genes for the related peptides of certain pituitary and placental hormones (D'Eustachio et al.), cluster of immune response and class I major histocompatibili-  
15 lity genes (Barsh et al.), clusters of genes for some interferons (Weiss et al.), genes for U1 and U2 RNAs (Erickson et al.), human metallothionein genes (Westin et al.), histone genes (Karin et al.), proline-rich salivary protein genes (Old et al.), genes for the constant regions of heavy or light  
20 chains of immunoglobins (Azen et al.), and alpha globin gene clusters (Tonegawa). The two globin gene clusters are especially relevant, because there are linked embryonic genes within these clusters which have important functions in early development (the zeta gene within the alpha cluster and the  
25 epsilon gene within the beta cluster). These would not have been readily detected by cross hybridization with the adult gene and therefore serve to illustrate the point that cross hybridization is not necessarily a good test of relatedness between clustered genes. Also relevant to this point are the  
30 problems encountered in attempting to isolate related lymphokines using cross-species homologies. Thus, the mouse IL-3 and GM-CSF genes share only limited homologies with their human counterparts and point to the possibility that growth factor

genes may have diverged very rapidly after tandemization (Wong et al., Yank et al.).

In general, genes which have arisen by duplication of a common ancestor prior to or in the early phases of the mammalian radiation have diverged to the extent that they cannot be readily detected by cross-hybridization of nucleic acid sequences. These observations indicate that, where divergent genes exist that arose by duplications of a common ancestor, moderate- to high- stringency hybridization approaches are probably not an adequate tool for their detection, whereas coding sequence comparisons between cDNAs may very well reveal functional similarities.

The cDNA amplification method described herein makes possible a method to isolate cDNAs that map to specific large DNA fragments. Isolation of these cDNAs then allows limited cDNA sequence analysis to determine whether they might encode interesting gene products and also provides probes to survey the normal tissue distribution of their corresponding transcripts. One application of this method has been directed to the isolation of new growth controlling genes.

The isolation of new growth-controlling genes would have considerable significance in investigating the normal physiological function of growth factors and in the production of sufficient quantities of material for the assessment of potential therapeutic applications. Figure 7 shows the regional assignments of potential growth factor/receptor genes within a subregion of chromosome 5. Five of these loci encode CSFs/interleukins. CSFs are a family of glycoproteins that are believed to be necessary for the growth and maturation of myeloid progenitor cells (Clark et al.). GM-CSF has been localized to 5q23-31 (Huebner et al., LeBeau et al., 1986). IL-3 (multi-CSF) is located within 5q23-31 (LeBeau et al., 1987). Initially, cDNA isolation efforts have been focused

predominately on the 5q23.3-q32 region of human chromosome 5 which contains eosinophil colony stimulating factor, also called T-cell replacing factor, (interleukin 5: IL-5) and the B-cell growth factor (IL-4) (Le Beau et al, VanLeeuwen et al).

5 A variety of sources can be used for the genomic DNA. A standard source is, of course, genomic DNA isolated from a tissue or a cell line. Hybrid cell lines also provide a convenient source of genomic starting material. Large DNAs of greater than 50 kb can be separated by a number of techniques,  
10 including pulse field gel electrophoresis (Carle et al., Smith et al., Chu et al.), to provide an appropriate starting material which has been enriched for the target region. Another convenient source of genomic starting material can be genomic DNA banks created in a cloning vector. There exist a number  
15 of vectors, known to one of ordinary skill in the art, which are useful for cloning large pieces of DNA: such useful cloning vectors include modifications of phage lambda and yeast artificial chromosomes. Lambda vectors (eg., Proto-clone™ lambda gt10, Promega) have been developed for creating  
20 genomic libraries which can tolerate relatively large insertions: for example, lambda gt10 can contain up to 7.6 kb inserts. Yeast artificial chromosomes (YAC) are in fact a preferred vector system since they provide the convenience of maintenance in a unicellular organism and have a capacity for  
25 very large inserts: up to several hundred kilobases (Burke et al.). In fact, the efficacy of the YAC system in the construction of a human genomic DNA bank has been demonstrated (Burke et al.).

A variety of methods to identify genomic DNA fragments  
30 of interest exists, such as hybridization with a labelled probe, but in many cases are extremely limited by the availability of sufficient quantities of target sequences for identification and subsequent molecular manipulations. The ampli-

fication method described in Section C is of great value in this regard since it provides a method to amplify a group of DNA molecules independent of prior knowledge of their nucleic acid sequence; an application of this method is presented in  
5 Example 5.

The overall scheme for purification of a genomic DNA fragment derived from the 5q23-31 region of human chromosome 5 is shown in Figure 8 and described in detail in Example 5. The source of starting DNA was the hamster/human hybrid cell  
10 line HHW105 (Overhauser et al.); this hybrid contains chromosome 5 as its only human material. Accordingly, use of this hybrid cell line provided an immediate enrichment of human chromosome 5 sequences. Initial pulse-field gel electrophoresis (Burke et al.) mapping using an IL-5 specific probe showed  
15 a NotI IL-5-containing fragment of 500 kilobases (kb) which was reduced by about 50 kb upon double-digestion with NruI; this fragment was chosen as the genomic DNA target fragment.

The HHW105 cell hybrid genomic DNA was cut with NotI; the resulting fragments were resolved on a CHEF gel (Smith et al., Chu et al.). Since CHEF gel tracks are straight rather  
20 than curved this gel system allowed the isolation of discrete size DNA fractions by horizontally slicing the gel. The IL-5 containing fraction was identified (page 9) by taking a small sample of each gel slice and subjecting the sample to the  
25 polymerase chain reaction assay (Scharf et al.) using IL-5 specific primers.

Our modifications of the polymerase chain reaction conditions make it unnecessary to purify the DNA away from the agarose in the gel slice sample, unlike previous methods which  
30 employ a phenol/chloroform extraction step to eliminate the agarose. The ability to eliminate the phenol/chloroform extraction step greatly facilitates performing the number of assays necessary to identify the genomic fragment of interest.

An approximation of the degree of purification that this method of genomic fragment isolation affords can be made by either estimating the amount of ethidium bromide fluorescence in the purified band vs. total, or by directly measuring the amount of DNA present in the size fraction by spectrofluorometry. Human chromosome 5 constitutes about 5%, by length, of the human genome and is present at one copy per genome in the diploid HHW105 thus constituting about 2.5% of all DNA present in the hybrid genome (Wasmuth personal communication). The following calculations can be made to assess the degree of purification required to isolate one specific large fragment of human DNA. The diploid genome size of human and hamster is  $6 \times 10^9$  base pairs (bp). Therefore, in HHW105 about  $1.5 \times 10^9$  bp of DNA will be contained in chromosome 5 (2.5% of  $6 \times 10^9$  bp). Assuming an average NotI fragment length of approximately 500 kb, chromosome 5 would contain 300 NotI fragments; accordingly, about a 300-fold purification would be required to isolate only one of these fragments. An 80-fold purification is readily obtained on one pulsed-field gel (Michiel et al) and two gel purification steps can raise this to at least 300-fold.

The IL-5-containing NotI genomic fragment was cut with NruI and this sample resolved on a second CHEF gel. The gel was sliced as above and an aliquot of each gel slice subjected to polymerase chain reaction amplification with IL-5 specific primers. These amplified samples were then size fractionated on a conventional agarose gel and the DNA molecules transferred to nitrocellulose. In this case the IL-5-containing fraction was identified by probing the nitrocellulose paper with a radiolabeled IL-5 specific probe (Figure 10).

The IL-5-containing NotI/NruI fragment and contaminating hamster fragments were further size fractionated, the A linkers (see Section B) ligated to the molecules, and the DNA

fragments amplified by the method described in Section C. These amplified fragments were then size fractionated on a conventional agarose gel (Figure 11) and transferred to a nitrocellulose filter. A variety of solid supports are available which will perform the same function as the nitrocellulose filter such as GeneScreen™ (DuPont-New England Nuclear) or the solid support described above in section 2A(ii).

Example 7 describes the selection of cDNA molecules which are homologous to the genomic regions isolated in Example 5. The cDNA molecules were selected from a T-cell cDNA library (Example 6). cDNA molecules can be amplified before and/or after selection. For example, vector sequences flanking the 5' and 3' end of the cDNA inserts can be used as end-terminal primers to amplify the cDNA inserts before any selection. The 5' ends of the primers can also contain sequences encoding restriction sites which will subsequently simplify insertion of the cDNA molecules into cloning/expression vectors.

Although a cDNA library was used in the IL-5 example (Example 7), the initial nucleic acid need not be a cDNA. It could, for example, be an RNA molecule where, after the initial selection by hybridization from a population of molecules, a first strand of homologous DNA is synthesized by a known procedure, such as random priming (Boehringer-Mannheim kit) or oligo-dT priming (Maniatis et al.), and second strand synthesis can be primed in the classic hairpin-primed fashion (Maniatis et al.).

Extensive washing of the filter, after hybridization, using relatively stringent conditions was performed to minimize non-specific binding. Since different linkers were used for amplification of the selected cDNAs versus the genomic sequences (the C and A primers, respectively, described in



Section B), any genomic sequences which come off the filter during the elution steps will not be amplified.

This selection technique is rapid and, to some extent, serves to equalize the abundance of selected cDNAs. This  
5 equalization occurs because the copy number of the genomic exon sequences on the filter should be roughly equivalent, despite differing abundances of the cDNAs homologous to each to these genomic sequences. Thus, when the cDNAs are in excess, a normalization should occur relative to the genomic  
10 sequences.

In the case of IL-5, Figure 12 shows the ethidium bromide staining of an agarose gel on which the amplified cDNA molecules, which were selected by the above described method, were resolved. Further, Figure 13 shows the hybridization of  
15 various probes with the amplified cDNAs. As can be seen from the figure, IL-5 specific sequences are present as expected. Further, a smear of hybridization corresponding to cDNA sequences homologous to total human DNA is also present suggesting a variety of cDNA molecules other than IL-5 have been  
20 isolated by this method.

The method applied to the isolation of cDNAs homologous to the IL-5 containing region of chromosome 5 has also been applied to the IL-3 (Figure 7) containing region of chromosome 5. As can be seen from Figure 14, the selection method was  
25 also efficacious for the isolation of cDNAs corresponding to the IL-3 region of chromosome 5. Using the genomic DNA clones we obtained from the IL-3 region we constructed a long-range restriction map of this region. Using a probe specific for GM-CSF we physically mapped GM-CSF to be within 10 kilobases of  
30 IL-3 (Figure 15). When cDNAs, selected on the basis of their hybridization to the IL-3 region genomic DNA probes, were tested for their hybridization to a GM-CSF specific probe a positive signal was obtained. This result further demonstrates

the ability of the instant method to identify cDNAs present in the same chromosomal region as a selected gene of interest (such as IL-3).

One important consideration in the above selection  
5 technique is the quenching of genomic repetitive sequences within the various cloned DNAs. Quenching can be achieved by one or both of the following treatments: 1) pre-hybridization of the filter; and 2) pre-hybridization of the cDNA library in  
10 solution before the filter selection step. Both of these approaches tie up the repetitive sequences and prevent them from acting as a homologous hybridization site in the selection step (e.g. the hybridization of the filter with the cDNA library). Effective quenching can be achieved in the following ways:

15 One approach to quenching is pre-hybridization to an intermediate  $C_t$  value (Litt et al., Sealey et al.) with total randomly sheared human genomic DNA.

Another approach to quenching genomic repeats is to use  
20 a chromosome-specific genomic DNA library from a chromosome other than the target chromosome for pre-hybridization: in the case of IL-5, a human chromosome other than human chromosome 5.

Finally, the most preferred approach to quenching genomic repeats is to use a library of isolated repetitive sequences representative of various abundance classes of repeat  
25 sequences. This set of repeats is isolated by screening a human genomic DNA library with total radiolabeled human DNA at a low plaque density. A mixture of plaques which show low, intermediate, and high density signals are picked and these  
30 sets of sequences can be used as the blocking reagent for pre-hybridization. Alternatively, such a library can be generated by shearing total human genomic DNA into short lengths, denaturing and reannealing to an intermediate  $C_t$  value and clon-

ing the repaired double-stranded molecules. In general, an initial blocking of the filter with sonicated herring sperm DNA is followed by pre-hybridization blocking using the repeat library.

5 To improve the representation of cDNAs corresponding to a given genomic region, a further step of hybridization selection is performed employing, for example: (1) a different starting combination of restriction enzymes used to generate the partially purified genomic size fraction; and/or (2) a  
10 different source of the genomic DNA region of interest (eg. a yeast artificial chromosome clone that contains the appropriate genomic region). The set of eluted cDNAs, from above, would be hybridized to these sequences. In this way the chances of re-electing contaminating cDNAs would be minimized and  
15 the chances of isolating only cDNAs from a specific region would be maximized.

Once cDNA molecules homologous to a selected chromosomal region have been isolated, the cDNAs are cloned (Protoclone™ lambda gt10) and plated at low plaque density. The phage DNAs  
20 are transferred to filters and hybridized (Woo) with a labeled repetitive sequence library or with total radiolabeled human DNA to distinguish between clones which contain repeat sequences from clones which contain unique sequences. DNA is then isolated from a representative sample of the unique  
25 sequence clones (500 clones in the case of IL-5) and the cDNA inserts isolated by EcoRI digestion. Aliquots of the cDNA inserts are spotted to nitrocellulose filters. Then the cDNA inserts are nick-translated and used to probe the nitrocellulose filters to identify duplicate clones (Southern); dupli-  
30 cates are then discarded.

The clones are then mapped to chromosome 5 using a mini-panel of the hybrid HHW105 DNA, hamster DNA, and human DNA. A large collection of various digest of these DNAs as Southern

blots have been prepared on nylon filters. Clones that map to chromosome 5 and are unique can be positioned on the map of the genomic DNA surrounding the IL-5 gene utilizing hybridization back to the large DNA fragments that were originally resolved by CHEF gel electrophoresis. Clones which are found to contain repeats will be pre-quenched or subcloned prior to further analysis.

This method provides a novel and efficient way to identify cDNAs corresponding to a contiguous region of genomic DNA when conventional probes are unavailable, and also the method provides a means to identify previously unknown gene products of potential research and therapeutic value.

#### 5. Identifying Homologous Viral Nucleic Acid Sequences

The methods of the instant invention are useful for the analysis of viral genome structure and function. Complete DNA copies of viral genomes can be used as probes or, alternately, selected fragments of the genome can be utilized. In the case of RNA viruses DNA copies of the genome can be synthesized, as by reverse transcriptase, or in some cases intermediate DNA forms of the virus may be isolated (Kashmiri et al.). Immobilized viral genomic DNA can be used to screen the nucleic acids of infected cells: for example, cDNA libraries may be made from the RNAs of infected cells and these cDNAs isolated on the basis of their sequence homology to the viral genomic DNA probe. The ability to amplify cDNA molecules independent of their specific sequence, that is, by use of the linkers (Figure 2) of the instant invention (e.g. Example 6), is helpful in the identification of normally low abundance messages. Further, the ability to amplify the cDNAs eluted from the viral genomic DNA probe provides another increase in the capability of the method to identify cDNAs having infrequent hybridizations.

The isolated cDNAs can then be mapped to the viral genome to assign coding regions. This method can also be used to isolate cellular proteins which are related to the viral genomic sequences.

5       The methods of the instant invention can also be applied to the analysis of the nucleic acid sequences of heterologous viruses. For example, a related virus can be used to help characterized a new virus and identify related coding sequences. This approach has been used to examine relationships  
10 between the bovine leukosis virus (BLV) and a newly isolated human-lymphoma associated virus (HuLAV, co-pending U.S. application 361,855, herein incorporated by reference). The structural proteins of HuLAV viruses demonstrate a high degree of homology to other known retroviruses; in particular, the  
15 seroreactivity of individuals seropositive for the 10C9 HuLAV, (the HuLAV carried in cell line 10C9) with BLV envelope protein suggests that the putative envelope protein of HuLAV 10C9 may share immunogenic epitopes with BLV.

A cDNA library prepared from HuLAV 10C9 particles (Example 9-B) was hybridized against a BLV genomic DNA probe  
20 bound to a nitrocellulose filter, the filter washed, the selected cDNAs eluted and cloned into lambda gt11. No cDNA molecules were selected which gave a positive signal when the lambda gt11 clones were immunologically screened with anti-BLV  
25 sera or anti-gp62 sera (gp62 is an envelope protein of another related retrovirus, HTLV-I; Kiyokawa et al).

However, HuLAV 10C9 cDNAs isolated by the methods of the instant invention, as described in Examples 9-10, yielded the following results: hundreds of lambda gt11 cDNA clones positive with anti-BLV; 15 clones positive with anti-p24 (p24 is  
30 an approximately 27,000 molecular weight HuLAV core protein) of which 4/15 cross-reacted with anti-BLV; and 4 positive

clones detected with anti-gp62 of which all four cross-reacted with anti-BLV.

Accordingly, use of the instant methods applied to the isolation of homologous nucleic acid sequences between the related but distinct viruses BLV and HuLAV resulted in ability to detect homologous nucleic acid sequences where standard technology had failed. This increased ability was evidenced by the immunological reactivities of the lambda gt11 expression products of the cloned selected cDNAs. Further, when the selected cDNAs were cloned in lambda gt10 and then screened with full length BLV probes, 76 positives were detected out of  $2 \times 10^4$  clones screened. These results illustrate the increased sensitivity which can be achieved by employing the methods of the instant invention.

15

#### 6. Enrichment of Selected Genomic DNA Fragments

One difficulty in mammalian genetics is having adequate amounts of starting material available for analysis. The present amplification method provides a straight-forward method to generate a library which is highly enriched in sequences derived from a selected genomic region independent of different sequence species. Genomic DNA fragments may be isolated from a contiguous genomic DNA section, eg. by size fractionation on a gel, fragmented to form fragments of less than about 2 kilobases in size, the linkers of the instant invention ligated onto the fragments, and the fragments amplified. As illustrated above for the IL-5 region of chromosomes, genomic sequences can be isolated for a region of interest. These genomic sequences can then be amplified independently of the specific fragment sequence using the linkers of the instant invention (Example 5-E) and then subcloned into an appropriate vector (eg. lambda gt10, Protoclone™, Promega). These cloned genomic sequences are then available for a variety of manipu-

lations such as use as probes (Examples 7-8, 9-10), especially in genetic linkage analysis, and the generation of long-range restriction maps. Such long-range restriction maps are constructed by restriction digestion of the various clones, size  
5 fractionation of their fragments, construction of restriction maps for each clone, and finally lining up overlapping regions between clones. Libraries generated in this way provide excellent representation of the genomic region of interest.

7. Identification of Disease-related Loci

10 The molecular identification and analysis of genes corresponding to disease-related loci has been a major challenge of molecular genetics. Genetic linkage data exists for a number of genetic diseases (e.g. familial schizophrenia (Sherrington et al.) and in some cases physical maps of the  
15 specific chromosomal regions are available (e.g. the cystic fibrosis locus has been mapped to an approximately 500 kb region (Poustka et al.)), growth factors involved in certain neoplasms such as the familial polyposis locus isolated to a small region of human chromosome 4 (Bodmer et al.). The  
20 method of the invention may be useful to aid in the identification of specific disease-related gene products. Figure 16 provides a general outline of approaches which may be taken leading to the identification of specific disease-related loci and their gene products.

25 Using the techniques outlined above for the isolation of genomic DNA fragments, long range restriction maps can be constructed of the genomic DNA region corresponding to a disease-related locus. A number of potential starting materials containing the genomic DNA of interest are available as  
30 discussed in section II-B, such as, yeast artificial chromosomes, hybrid cell lines, and DNA fragments obtained by chromosome jumping (Collins et al.). Then, by application of the techniques used for the isolation of the genomic DNA in

the region of IL-5, known genetic markers for the regions of disease related loci can likewise be used to isolate genomic DNAs from these regions and then these selected genomic DNAs can be used as probes to select corresponding cDNA molecules  
5 from normal and disease-state samples.

Candidate cDNAs from the region of interest can then be isolated and compared by a variety of methods to determine differences between normal and disease state samples. One approach is the use of the subtraction method described above  
10 to detect cDNAs which are present in one sample, but not present in another. Another approach is to directly select a small set of cDNAs using the technology described above. To identify specific mutations in the isolated cDNAs a number of systems can be used including the following: (1) the use of  
15 modified denaturing gradient polyacrylamide gels to identify single base pair alterations, and (2) the use of incorporated biotinylated nucleotides to mark the positions of mutations determined by comparison to wild type DNAs. Also, partial DNA sequence analysis can be used to facilitate matching and  
20 comparisons of candidate cDNAs obtained from different samples, as well as to eliminate candidates based on their homology to known sequences present in sequence data bases (e.g. N.I.H. GeneBank).

The following examples illustrate the method of fragment  
25 amplification and isolation, and of the cDNA isolations described above, but are in no way intended to limit the scope of the method or its application.

#### Example 1

##### 30 General Procedures and Materials

Plasmid pIAN13 was obtained from Dr. Brian Seed (MIT).  
E. coli strain KM392, a suppressor positive strain, was ob-



tained from Dr. Kevin Moore, DNAX, Palo Alto, CA. E. coli strain LG75, a suppressor minus strain, was obtained from Dr. Brian Seed (MIT). Lambda phage Ch21a was obtained from Dr. Frederick Blattner (University of Wisconsin), and a NotI site  
5 was inserted between the EcoRI and HindIII sites (Ch21aLJ), by replacing the 2 kb EcoRI/ HindIII fragment with an oligonucleotide containing a NotI site. E. coli strain MC1061 harboring the P3 plasmid was obtained from Dr. Brian Seed (MIT).

Terminal transferase (calf thymus), alkaline phosphatase  
10 (calf intestine), polynucleotide kinase, E. coli DNA polymerase I (Klenow fragment), and S1 nuclease were obtained from Boehringer Mannheim Biochemicals (Indianapolis, IN); phiX174 fragments, produced by digesting the phiX174 phage to completion with HaeIII, were obtained from BRL Laboratories (Bethesda, MD).  
15

SmaI, EcoRI, NotI, T4 DNA ligase and T4 DNA polymerase were obtained from New England Biolabs (Beverly, MA); and streptavidin agarose, from Bethesda Research Labs (Bethesda, MD). Low-gelling temperature agarose (Sea Plaque) was ob-  
20 tained from FMC (Rockland, ME). Nitrocellulose filters were obtained from Schleicher and Schuell.

Synthetic oligonucleotide linkers and primers were prepared using commercially available automated oligonucleotide synthesizers. Alternatively, custom designed synthetic  
25 oligonucleotides may be purchased, for example, from Synthetic Genetics (San Diego, CA). cDNA synthesis kit and random priming labeling kits were obtained from Boehringer-Mannheim Biochemical (BMB, Indianapolis, IN).

Kinasing of single strands prior to annealing or for  
30 labeling is achieved using an excess, e.g., approximately 10 units of polynucleotide kinase to 1 nmole substrate in the presence of 50 mM Tris, pH 7.6, 10 mM MgCl<sub>2</sub>, 5 mM dithioth-

reitol, 1-2 mM ATP, 1.7 pmoles gamma<sup>32</sup>P-ATP (2.9 mCi/mmole), 0.1 mM spermidine, 0.1 mM EDTA.

Site-specific DNA cleavage is performed by treating with the suitable restriction enzyme (or enzymes) under conditions which are generally understood in the art, and the particulars of which are specified by the manufacturer of these commercially available restriction enzymes (see, e.g., New England Biolabs, Product Catalog). In general, about 1 µl of plasmid or DNA sequence is cleaved by one unit of enzyme in about 20 µl of buffer solution after 1 hr digestion at 37°C; in the examples herein, typically, an excess of restriction enzyme is used to insure complete digestion of the DNA substrate.

Incubation times of about one hour to two hours at about 37°C are workable, although variations can be easily tolerated. After each incubation, protein is removed by extraction with phenol/chloroform, and may be followed by ether extraction, and the nucleic acid recovered from aqueous fractions by precipitation with ethanol (70%). If desired, size separation of the cleaved fragments may be performed by polyacrylamide gel or agarose gel electrophoresis using standard techniques. A general description of size separations is found in Methods in Enzymology (1980) 65:499-560.

Restriction cleaved fragments may be blunt-ended by treating with the large fragment of E. coli DNA polymerase I (Klenow reagent) in the presence of the four deoxynucleotide triphosphates (dNTPs) using incubation times of about 15 to 25 min at 20°C to 25°C in 50 mM Tris pH 7.6, 50 mM NaCl, 6 mM MgCl<sub>2</sub>, 6 mM DTT and 0.1-1.0 mM dNTPs. The Klenow fragment fills in at 5' single-stranded overhangs in the presence of the four nucleotides. If desired, selective repair can be performed by supplying only one of the, or selected, dNTPs within the limitations dictated by the nature of the overhang.

After treatment with Klenow reagent, the mixture is extracted with phenol/chloroform and ethanol precipitated. Treatment under appropriate conditions with S1 nuclease results in hydrolysis of any single-stranded portions of DNA. In particular, the nicking of 5' hairpins formed on synthesis of cDNA is achieved.

Ligations are performed in 15-50  $\mu$ l volumes under the following standard conditions and temperatures: for example, 20 mM Tris-Cl pH 7.5, 10 mM  $MgCl_2$ , 10 mM DTT, 33 mg/ml BSA, 10 mM-50 mM NaCl, and either 40 mM ATP, 0.01-0.02 (Weiss) units T4 DNA ligase at 14°C (for "sticky end" ligation) or 1 mM ATP, 0.3-0.6 (Weiss) units T4 DNA ligase at 14°C (for "blunt end" ligation). Intermolecular "sticky end" ligations are usually performed at 33-100 mg/ml total DNA concentrations (5-100 nM total end concentration). Intermolecular blunt end ligations are performed at 1 mM total ends concentration.

In vector constructions employing "vector fragments", the vector fragment is commonly treated with bacterial alkaline phosphatase (BAP) or calf intestinal alkaline phosphatase (CIP) in order to remove the 5' phosphate and prevent self-ligation of the vector. Digestions are conducted at pH 8 in approximately 10 mM TrisHCl, 1 mM EDTA using about 1 unit per mg of BAP at 60°C for one hour or 1 unit or CIP per mg of vector at 37°C for about one hour. In order to recover the nucleic acid fragments, the preparation is extracted with phenol/ chloroform and ethanol precipitated. Alternatively, religation can be prevented in vectors which have been double digested by additional restriction enzyme digestion and separation of the unwanted fragments.

## Example 2

Amplifying Duplex DNA Fragments

## A. Modification and Amplification of pIAN13 plasmid

Plasmid pIAN13 carries a supF suppressor gene capable of  
5 suppressing an amber mutation. The plasmid was linearized by  
digestion with SmaI and ligated to the linker whose sequence  
is shown at A in Figure 2, at a 1:100 molar ratio in the  
presence of 0.3-0.6 Weiss units of T4 DNA ligase. The linker  
10 sticky ends were filled in with Klenow fragment, conventional-  
ly, and the mixture was treated with NruI to cleave linker  
dimers. SmaI linkers were attached to the blunt-ends of the  
plasmid fragments under ligation conditions like those above.  
After treatment with SmaI to remove redundant SmaI linkers,  
the fragments were recircularized. This was done at a frag-  
15 ment concentration of about 1 µg/ml, in the presence of 4,000  
units/ml DNA ligase, at about 14°C for 18 hours. The result-  
ing plasmids were thus modified to contain an A-sequence  
linker on either side of the original SmaI site.

E. coli strain MC1061 harboring the P3 plasmid was  
20 transformed with the recircularized plasmid, and selection for  
successful plasmids was made on the basis of bacterial growth  
in the presence of kanamycin, ampicillin, and tetracycline.  
Plasmids were isolated from the resistant bacterial colonies,  
and cut with SmaI to produce blunt-end, linearized fragments  
25 having the desired A-sequence end linkers.

To 100 µl of 10 mM Tris-Cl buffer, pH 8.3, containing  
1.5 mM MgCl<sub>2</sub>, (Buffer A) was added 1 x 10<sup>-3</sup> µg of the linearized  
plasmid, 2 µM of a primer having the sequence  
d(5'-GGAATTCGCGGCCGCTCG-3'), 200 µM each of dATP, dCTP, dGTP,  
30 and dTTP, and 5 units of Thermus aquaticus DNA polymerase (Taq  
polymerase). The reaction mixture was heated to 94°C for 1  
minute for denaturation, allowed to cool to 50°C for 2 min  
for primer annealing, and then heated to 72°C for 5-12 min to

allow for primer extension by Taq polymerase. The replication reaction, involving successive heating, cooling, and polymerase reaction, was repeated an additional 25 times with the aid of a Perkin-Elmer Cetus DNA thermal cycler.

5

#### B. Amplification of phiX174 Fragments

To a 50 µl solution containing 100 ng of the HaeIII blunt-end fragments was added 300 ng of the linker whose sequence is shown at C in Figure 2, with ligation of the  
10 linker in the presence of DNA ligase being carried out as above. The mixture was treated with NruI to cleave linker dimers.

To 100 µl of Buffer A was added 1 ng of the phiX174/HaeIII fragments, 2 µM of a primer having the sequence  
15 d(5'-CACGCGTGAATTCTCG-3'), 200 µM each of dATP, dCTP, dGTP, and dTTP and 5 units of Taq DNA polymerase. The reaction mixture was heated to 94°C for 1 minute for denaturation, allowed to cool to 50°C for 2 min for primer annealing, and then heated to 72°C for 5-12 min to allow for primer extension  
20 by Taq polymerase. The replication reaction steps were repeated 25 times, as above. The amplified fragments are referred to below as phiX/linker-C fragments.

A second preparation of phiX174 fragments was amplified by similar methods, but using the sequence-A linker employed  
25 in the piAN13 amplification. This preparation is referred to below as phiX/linker-A fragments.

### Example 3

#### Selection for piAN13 Sequences

##### 30 A. Biotinylation of Amplified phiX174 Fragments

The amplified phiX174/linker-C fragment produced in Example 2-B are treated with EcoRI to yield 5' protruding ends

having an AATT sequence overhang. The fragments are biotinylated by Klenow fill-in reaction according to known procedures (Maniatis, pp. 113). Briefly, the reaction mixture consists of 100 µg/ml fragments in 100 mM potassium phosphate (pH 7.2),  
5 2 mM CoCl<sub>2</sub>, 0.2 mM DTT, 40 µM Bio-7-dATP, 100 µM dTTP, and 100 units/ml Klenow fragment. After incubation at 25°C for 45 minutes, the reaction is terminated by heating at 65°C for 10 minutes.

10 B. Hybridization Reaction

Amplified pIAN13 vector is mixed at various mole ratios with the phiX/Linker-A fragments produced in Example 2-B. Mixtures contained weight ratios of 1:100, 1:30 and 1:3 pIAN13 to phiX/Linker-A fragments are prepared. Each mixture is then  
15 mixed with a 10-fold molar excess of biotinylated phiX/Linker-C fragments prepared as above, in an annealing buffer containing 2M guanidinium isothiocyanate (pH 5.8). The mixtures are heated to 65°C for 5 minutes, then allowed to cool to 25°C and hybridized for a 12-hour period.

20

C. Separation by Affinity Chromatography

A 1 ml siliconized syringe plugged with siliconized glass wool is packed with 0.5 ml streptavidin-agarose and washed with 10mM Tris-Cl, 0.5 M NaCl, pH 7.0, containing 1 mM  
25 EDTA (elution buffer). A portion of each of the three hybridization mixtures from above is EtOH precipitated (70%), resuspended in 10 mM Tris HCl, 1 mM EDTA pH 7.0, and then loaded onto the column which is washed with several volumes of column elution buffer.

30 For each mixture, the eluted fractions from the streptavidin column are combined and the DNA concentrated by ethanol precipitation. The DNA is resuspended in Buffer A to a

concentration of 0.01  $\mu\text{g/ml}$ , and mixed with 2  $\mu\text{M}$  of the above primer having the sequence d(5'-GGAATTCCGGCCGCTCG-3'), and 200  $\mu\text{M}$  each of dATP, dCTP, dGTP, and dTTP, and 5 units Taq polymerase. The reaction mixture is denatured at 94°C for 1 minute, cooled to 50°C for primer annealing, and primer extension performed at 72°C for 5-12 minutes.

The fragments from the each of the three eluted mixtures above are cut with NotI to produce NotI sticky ends in the linker regions of the fragments, as can be appreciated from Figure 2A. The fragments (13  $\mu\text{g/ml}$ ) are mixed with 60  $\mu\text{g/ml}$  NotI-digested phage Ch21aLJ, and inserted at the NotI site in the phage in the presence of T4 DNA ligase. After encapsulation, the phage are used to infect E. coli strain KM392, a suppressor-positive strain, and E. coli strain LG75, a suppressor-minus strain. The number of plaques produced in the KM392 strain provides a measure of total number of recombinant phage, since this strain is able to suppress the phage amber mutation, and thus allow phage expression of its A and B genes. In the LG75 strain, by contrast, only phage which have incorporated the piAN13 fragment produce plaques, since the sup F gene is required for growth of phage in suppressor minus bacterial strains. The ratio of plaques on the two host strains thus indicates the ratio of piAN13 fragments in the fragment mixtures. As a control, the ratio of fragments in the three mixtures before hybridization and streptavidin separation is also determined.

#### Example 4

##### Screening Amplified cDNA Fragment Mixtures

###### 30 A. Bile-derived cDNAs

Two cynomolgus monkeys were intravenously injected with a 10% suspension of stool positive for ET-NANB obtained from a

human volunteer. It was determined the stool was positive for ET-NANB by binding of viruslike particles (VLPs) in the stool to immune serum from a known ET-NANB patient. The monkeys developed elevated levels of alanine aminotransferase (ALT) between 24-36 days after inoculation, and excreted 27-30 nm VLPs in their stools in the pre-acute phase of infection. The animals became seropositive for VLPs in the inoculum.

The bile duct of each infected animal was cannulated and about 1 cc of bile was collected. RNA was extracted from the bile by hot phenol extraction, using a standard RNA isolation procedure. Double-strand cDNA was formed from the isolated RNA by a random primer for first-strand generation, using a cDNA synthesis kit obtained from Boehringer-Mannheim (Indianapolis, IN).

A cDNA fragment mixture from non-infected cynomolgus monkey bile was prepared similarly.

#### B. Stool-Derived cDNAs

cDNA fragments obtained from human stool samples were prepared as follows: a 10% suspension of stool samples from healthy or ET-NANB-infected individuals was layered over a 30% sucrose density gradient cushion, and centrifuged at 25 kg for 6 hr in an SW27 rotor, at 15°C. The pelleted material contained 27-32 nm VLP particles characteristic of ET-NANB infection in the infected-stool sample. RNA was isolated from the sucrose gradient pellets in both the infected and noninfected samples, and the isolated RNA was used to produce cDNA fragments as described in Example 2.

#### C. Amplification of the cDNA Fragment Mixtures

The cDNA fragment mixtures from infected and non-infected bile source, and from infected and noninfected human-stool source were each amplified by linker/ primer replica-



tion, according to the method of the invention. Briefly, the fragments in each sample were blunt-ended with DNA Pol I then extracted with phenol/chloroform and precipitated with ethanol. The blunt-ended material was ligated with linker A shown  
5 in Figure 2, under conditions like those described in Example 2, and the mixtures were digested with NruI to remove linker dimers. Each mixture was brought to a final concentration of 0.01 µg/ml in Buffer A, and to this was added 2 µM of a primer having the sequence d(5'-GGAATTCGCGGCCGCTCG-3'), and 200 µM  
10 each of dATP, dCTP, dGTP, and dTTP. The reaction mixture was heated to 94°C for 1 minute for denaturation and allowed to cool to 50°C for 2 min for primer annealing, and then heated to 72°C for 5-12 min to allow for primer extension by Taq polymerase. The replication reaction, involving successive  
15 heating, cooling, and polymerase reaction, was repeated an additional 25 times with the aid of a Perkin Elmer Cetus DNA thermal cycler.

#### D. Screening Amplified cDNA Mixtures

20 The four amplified cDNA fragment mixtures from above were fractionated by agarose gel electrophoresis, using a 2% agarose matrix. After transfer of the DNA fragments from the agarose gels to nitrocellulose paper, the filters were hybridized to a <sup>32</sup>P random-primed 1.33 kb fragment associated with  
25 ET-NANB viral agent. The derivation and sequence of this fragment, which is carried in a pTZ-KF1(ET1.1) plasmid identified by ATCC No. 67717. The cloned fragment was obtained by digesting the pTZ-KF1(ET1.1) plasmid with EcoRI, (ii) isolating the released 1.33 kb ET-NANB fragment, and (iii) random  
30 priming the isolated fragment.

The probe hybridization with the fractionated cDNA mixtures was performed by conventional Southern blotting

methods (Maniatis, pp. 382-389). Figure 6 shows the hybridization pattern obtained with cDNAs from infected and non-infected bile sources. As seen, the ET-NANB probe hybridized with amplified fragments obtained from the infected source, but was non-homologous to sequences obtained from the non-infected source. Similar results were obtained with amplified fragments obtained from infected and non-infected human stool sources. This procedure demonstrates the utility of the described invention since hybridization analysis of an equivalent amount of cDNA prior to amplification would not have yielded detectable hybridization signal.

#### Example 5

##### Isolation of Genomic DNA from the region adjacent the gene encoding IL-5

##### A. Preparation of Large DNA fragments

The hybrid cell line HHW105 (Overhauser et al.) was used as the source of genomic DNA. This cell line contains all of human chromosome 5. The human chromosome is retained by selection for the leucine aminoacyl tRNA synthetase gene on human chromosome 5 which complements a corresponding defective hamster gene. This hybrid cell line was employed to obtain an immediate enrichment of chromosome 5 sequences.

The gene encoding IL-5 is located in the 5q31.3 region of chromosome 5 (LeBeau et al). Genomic DNA was isolated from cell line HHW105 (Maniatis et al, Smith et al) and digested with a variety of restriction endonucleases. The resulting genomic DNA fragments were size fractionated by Pulse-Field Gel Electrophoresis (Carle and Olson). The DNA fragments in this gel were then transferred to nitrocellulose (Southern) and the nitrocellulose filter hybridized with a random-primed IL-5 cDNA fragment (Azume et al). A NotI fragment which hybridized with the IL-5 probe was identified. This 500 kilobase (kb)

NotI fragment was reduced by about 50 kb upon double digestion with NruI; this NotI/NruI fragment was chosen as a target region for purification and subsequent use in homologous cDNA selection. The purification of this fragment is outlined in Figure 8 and described below.

5 X 10<sup>9</sup> cells were harvested and washed once in Dulbecco's Phosphate Buffered Saline (PBS). The cells were resuspended in PBS at 1 X 10<sup>6</sup> cells/ml to which was added an equal volume of L-buffer (50 mM NaCl, 50 mM EDTA, pH=8) and 1% low gelling temperature (LGT) agarose (FMC Bioproducts) held at 50°C. Proteinase K was then added to a final concentration of 0.5 mg/ml. The mixture was gently pipetted into block formers and the blocks allowed to solidify. These blocks were then placed in 50 ml conical tubes, totally immersed in 40 mls of 0.5 M EDTA, pH=9, 1% sarkosyl, and 200 µg/ml proteinase K, and then incubated overnight at 50°C. The 40 mls of solution was then replaced with L-buffer containing PMSF at a final concentration of 1mM and incubated for 2 hours at room temperature; this wash was repeated. The blocks were then washed twice, with 2 hour incubations for each wash, in TE (10 mM Tris HCl, pH=8, 1 mM EDTA) at room temperature.

B. The First CHEF gel.

A 1% LGT agarose, 0.5 X TBE (5 X TBE stock solution, per liter: 54 g Tris Base, 27.5 g boric acid, and 20 mls 0.05 M EDTA (pH=8)) gel was poured containing 4 conventional wells and one long preparative well. Sixteen blocks of HHW105 DNA were equilibrated in 1 X NotI reaction buffer (New England Biolabs) for two incubations of one hour each. The equilibrated blocks were incubated with NotI at 37°C, for four hours, under the following digestion conditions: 16 blocks in 15 ml corex tube, 200 µl 10 X NotI buffer, 20 µl 10 mg/ml

acetylated bovine serum albumin (BSA), 200  $\mu$ l NotI enzyme (10 units per  $\mu$ l), 1560  $\mu$ l water. The blocks were then treated for 2 hours in 1 mM PMSF. One individual block, lambda DNA ladder size markers (c185757, FMC Bioproducts) and Schizosac-  
5 charomyces pombe chromosomal DNA (972, FMC Bioproducts) were loaded in the conventional wells. The remaining blocks were heated to 65°C, melted and loaded in the molten state into the preparative well. The gel was run under standard CHEF gel conditions (Smith et al., Chu et al.) for 48 hours at 200 V  
10 with 82 second switching intervals. The portion of the gel containing the 4 conventional lanes was then cut away from the preparative well and stained with ethidium bromide (see Figure 9-A, First CHEF gel). Since CHEF gel tracks are straight rather than curved, DNA fractions covering size ranges across  
15 the gel could be collected. The preparative portion of the gel was sliced, using a glass cover slip, into 1-2 mm slices covering the size range from approximately 1000 kb to 150 kb for a total of 30 slices down the gel.

20 C. Identification of the IL-5 containing fraction

Identification of the IL-5 containing fraction was achieved by sequence-specific amplification using IL-5 specific primers (IL-5-5' = 5'-GGGAATTCATGAGGATGCTTCTGCATTTG-3'; IL-5-3' = 5'-GGAAGCTTTCAACTTTCTATTATCCACTCGGTGTTTCATTAC-3').  
25 DNA was extracted from a 1 mm by 2 mm portion of each slice of the gel; the rest of each slice was stored until the fraction containing the IL-5-homologous fragment was identified. The 1 X 2 mm sample was placed in a microfuge tube, heated to 65°C until the agarose was molten (approximately 5 minutes), mixed  
30 gently and a 20  $\mu$ l aliquot was placed on a piece of Parafilm™ and allowed to solidify to an agarose bead. The agarose bead was placed in 200  $\mu$ l of 1 X polymerase chain reaction buffer

(Perkin-Elmer Cetus Gene Amplification Kit) and equilibrated at room temperature for 30 minutes. The buffer was then removed and the following standard reaction mixture added to the agarose bead: 16  $\mu$ l 1.25 mM dNTPs, 10  $\mu$ l 10 X polymerase chain reaction buffer, 5  $\mu$ l 20  $\mu$ M IL-5-3' primers, 5  $\mu$ l 20  $\mu$ M IL-5-5', 63.5  $\mu$ l water. The mixture was heated to 65°C until the agarose was molten and then 0.5  $\mu$ l of Taq I DNA polymerase was added. The mixture was briefly vortexed, a thin layer of mineral oil placed on top of the reaction mixture, and the amplification reaction run under the following conditions: denature for 1 minute at 94°C, anneal for 1 minute at 50°C, extend for 2 minutes at 72°C, repeat for 25 cycles. Following the amplification, a conventional agarose gel was run to determine the fraction containing IL-5 (see Figure 9, polymerase chain reaction products); in parallel three control polymerase chain reaction reactions were run on each gel: human DNA with IL-5 primers, hamster DNA with IL-5 primers, and IL-5 primers alone. An example of a single CHEF gel track of NotI digested HHW105 genomic DNA and of the polymerase chain reaction products is shown in Figure 9. As seen in this figure, a single DNA sample contained the majority of the IL-5 fragment; this polymerase chain reaction product comigrated with the human DNA control polymerase chain reaction product. The hamster DNA and primers alone showed no polymerase chain reaction products.

#### D. The Second CHEF Gel

In order to achieve a further purification of the relevant fragment from the contaminating hamster fragments, the DNA in the slice corresponding to the IL-5-positive polymerase chain reaction product was subjected to digestion with the NruI restriction endonuclease, and size fractionated on a CHEF

gel (as described above for NotI in Section B). The gel was sliced as above and a small aliquot of DNA obtained from each slice was subjected to polymerase chain reaction amplification with the IL-5-specific primers to determine the positive fractions. The polymerase chain reaction amplification products were resolved on a conventional agarose gel, the DNA molecules transferred to nitrocellulose, and, in this case, the IL-5-specific fraction was identified by hybridization with a radiolabeled IL-5 cDNA probe. The resulting autoradiogram is shown in Figure 10.

E. Amplification of the IL-5 positive fraction by sequence-independent amplification.

In order to have sufficient quantities of these genomic DNA fragments for use in cDNA library screening and for the construction of a library of these genomic fragments, the genomic DNA fragments were amplified using the sequence-independent single primer amplification (SISPA) method described in Section C. To generate the blunt-ended molecules necessary for the linker addition, the DNA from the relevant size fraction which was identified in the second CHEF gel was digested with EcoRV, which generates blunt end cuts in the DNA at a hexanucleotide recognition site. The digestion products were fractionated on a conventional 0.8% agarose gel with size markers in parallel. The gel was rotated through 90° and the digested DNA was electroeluted onto a charged membrane filter (NA45, Schleicher and Schuell). The filter was cut into ten strips, five corresponding to the positions of DNAs of 2 kb or less in size, and five corresponding to DNAs above 2 kb in size. The DNAs were eluted from the filter in very small volumes (elution buffer: 1 M NaCl, 50 mM Arginine for one hour at 65°C), phenol/chloroform extracted, and precipitated in the presence of carrier t-RNA. DNAs above 2 kb in length were

subjected to one more set of restriction endonuclease digestions using HaeIII and RsaI; both of these enzymes generate blunt ends and recognize four base pair sites. In this way the larger DNAs were reduced to lengths that could be more efficiently amplified. Each of the ten fractions were ligated using T4 DNA ligase (New England Biolabs) to the A linkers (Figure 2-A) and were then cut with NruI to eliminate any linker dimers which had been formed in the ligation reaction. The volume of the restriction endonuclease digestion reactions were increased to 100  $\mu$ l. The fractions were then passed over Sephadex<sup>™</sup> G-50 (Maniatis et al., page 466).

The genomic fragments were then amplified, as described in Section C and Example 2 using Taq DNA polymerase I (Perkin-Elmer Cetus) and the A primer under the following conditions: 25  $\mu$ l of the linkered fraction, 5  $\mu$ l of 20  $\mu$ M A primer, 10  $\mu$ l 10 X amplification reaction buffer (100 mM Tris HCl, pH=8.3, 500 mM KCl, 15 mM MgCl<sub>2</sub>, 0.1% (w/v) gelatin), 16  $\mu$ l 1.25 mM dNTPs, 43.5  $\mu$ l water, 0.5  $\mu$ l Taq I DNA polymerase. The reaction mixture was cycled 25 times for the following periods: denature at 94°C for 1 minute, anneal at 50°C for 1 minute, and extension at 72°C for 2 minutes. An aliquot of each amplified fraction was separated on a 1% agarose gel which is shown in Figure 11. A smear of amplified products was seen. Carrier t-RNA treated in parallel does not amplify. The furthest most right hand lane of Figure 11 is the size standard lambda HindIII. The DNA in this gel (the final SISPA gel) was transferred to a GeneScreen<sup>™</sup> filter (DuPont/New England Nuclear), the DNA cross-linked to the filter by UV, and the filter pre-hybridized overnight at 65°C in GeneScreen<sup>™</sup> buffer (Church and Gilbert) with the addition of 100  $\mu$ g/ml of sheared herring sperm DNA (the final SISPA filter, see Example 7).

An aliquot of the amplified genomic fragments (not bound to nitrocellulose) was used for the construction of a fragment library.

5

#### Example 6

##### Preparation of the T-cell cDNA library

T-cells were separated from whole blood by standard methods and the leu3+ fraction specifically selected. The leu3+ T-cells were induced with PMA (phorbol myristate acetate) by standard procedures. RNA was extracted from the T-cells 6 hours post-induction with PMA, and the poly-A<sup>+</sup> RNA fraction was purified from T-cell total RNA by affinity chromatography on an oligo-dT cellulose column. A leu 3+ T-cell cDNA library was prepared essentially as described by Maniatis et al (pages 229-246). Synthesis of the first strand was achieved using an oligo-dT primer and reverse transcriptase. Second strand cDNA synthesis was performed by a combination of the Klenow fragment of DNA polymerase I, followed by treatment with reverse transcriptase. Hairpin structures were resolved with S<sub>1</sub> nuclease. EcoRI linkers were blunt end ligated onto the newly synthesized duplex cDNA molecules and these molecules were then ligated into the lambda gt10 vector for cloning purposes (Protoclone™ lambda gt10, Promega).

A 20 µg aliquot of the T-cell DNA library DNA was digested with EcoRI to liberate the cDNA inserts from the phage arms (approximately 4% of the total weight is insert cDNA). The EcoRI ends were then converted to blunt ends by filling in with dATP and TTP, using the Klenow fragment of DNA polymerase I (Maniatis et al., page 113). The use of only dATP and TTP repairs only the EcoRI sites but not the cos sites of lambda, thus avoiding the potential problem of creating blunt end termini on both ends of the contaminating phage vector DNA. The C linker (Figure 2-C) was then ligated to this total cDNA



library and the ligation mixture digested with NruI to remove dimers. The volume of the reaction mixture was then increased to 1.0 ml with GeneScreen™ buffer, sheared lambda gt10 DNA added to a final concentration of 100 µg/ml, and the mixture  
5 incubated at 65°C for 90 minutes; this mixture is the probe used in Example 7.

In addition, this preparation of the cDNA library may be pre-hybridized to randomly-sheared total human genomic DNA, before hybridization to the final SISPA filter, to pre-block  
10 repetitive sequences present in the cDNA library and thus reduce the background/signal noise (see Section ii-b(III)). The cDNA library can be amplified using the C-linker sequence as an end-terminal primer for amplification (Example 7). Alternatively, before the cDNA molecules are liberated from  
15 the lambda gt10 vector, common vector sequences flanking the 5' and 3' end of the cDNA inserts can be used as end-terminal primers to amplify the cDNA inserts.

#### Example 7

#### 20 Hybridization of the T-cell cDNA library to the Chromosome 5 derived genomic DNA

##### A. Hybridization and Washing

The T-cell cDNA library probe constructed in Example 6 was added to the final SISPA filter prepared in Example 5 in a  
25 total volume of 10. mls of GeneScreen™ buffer. The hybridization was carried out for 48 hours at 65°C. The filter was then extensively washed under the following conditions:

1. Two washes with 1 X SSC / 0.1% SDS at room temperature for 15 minutes each.
- 30 2. Two washes with 1 X SSC / 0.1% SDS at 65°C for 15 minutes each.
3. Two washes with 0.1 X SSC / 0.1% SDS at 65°C for 15 minutes each.

4. A final wash with 0.1 X SSC / 0.1% SDS at 65°C for 20 hours.

B. Elution of the selected cDNAs from the filter

5 After the final wash, the selected cDNAs were eluted from the filter by treatment with 5 mls of 50 mM NaOH at room temperature. This eluent was neutralized to pH=8.0 with approximately 2 mls of 1.0 M Tris HCl. The cDNAs were precipitated with NaCl, yeast RNA, and ethanol, and then resuspended  
10 in 100 µl of water. The resuspended cDNAs were then passed over a Sephadex™ G-50 column.

C. Amplification and Analysis of the selected cDNAs

The cDNAs present in the eluent were then amplified  
15 using the C primer and the following conditions: 30 µl eluted material, 10 µl 10 X amplification reaction buffer, 16 µl 1.25 mM dNTPs, 5 µl 20 µM C primer, 45 µl water, 0.5 µl Taq I DNA polymerase. The amplification reaction was carried out for 25 cycles as follows: denature at 94°C for 1 minute, anneal at  
20 45°C for 1 minute, and extend at 72°C for 2 minutes. The amplification products were run on a conventional agarose gel to examine the products (Figure 12). The DNA in the gel was transferred to a nitrocellulose filter and the filter hybridized with 6 radiolabeled probes: actin, hamster genomic DNA,  
25 human genomic DNA, IL-3 specific (Yang et al.), IL-5 specific (Azume et al.), and human mitochondrial DNA specific (CF14) (Figure 13).

Example 8Selection of cDNAs corresponding to the genomic region  
containing IL-3 and GMCSF

The methods applied to the analysis of IL-5 have also  
5 been applied to IL-3. Specific probes for IL-3 (Yang et al.)  
and GM-CSF (Schrader et al.) are known. Using an IL-3 specific  
probe the IL-3 containing region of chromosome 5 was isolated  
as described above; a long-range restriction map of this  
region is shown in Figure 15. By probing isolated genomic  
10 clones with labelled probes specific for IL-3 and GM-CSF we  
showed a physical linkage between these loci of less than 10  
kilobases. T-cell cDNAs were then selected on the basis of  
their homology to the selected IL-3 containing region genomic  
DNA region. Figure 14 shows the hybridization of these selec-  
15 ted cDNAs with probes specific for IL-3, GM-CSF, and actin,  
and also a total human genomic probe. Specific hybridization  
to IL-3 and GM-CSF was clearly visible. The lack of hybridiza-  
tion to actin underscores that the hybridization to IL-3 and  
GM-CSF was specific and the lack of hybridization to total  
20 human genomic DNA indicates the absence of large numbers of  
repetitive sequence DNA in the selected cDNAs.

Example 9Isolation of cDNAs derived from 10C9 mRNAs which are  
25 homologous to the Bovine Leukosis Virus genomeA. Preparation of the Genomic Bovine Leukosis Virus DNA

The cloning and sequencing of the bovine leukosis virus  
(BLV) genome have been previously described (Kashmiri et al.,  
Sagata et al.). Full length molecular DNA clones of BLV cor-  
30 responding to the RNA genome of the virus were obtained from  
J.F. Ferrer, University of Pennsylvania, School of Veterinary  
Science. 100 nanograms of BLV DNA was heat denatured to yield

single-strands and spotted to a 0.45 micron nitrocellulose filter (Schleicher and Schuell). The filter was then baked for 2 hours at 75° under vacuum to fix the DNA to the filter (Ricciardi et al.). The following solution was used for pre-  
5 hybridization of the filter: 50 mM Tris HCl, pH=7.5; 1 mM EDTA; 0.5% SDS; 3 X SSC (20 X SSC stock solution = 175.3 g NaCl, 88.2 g sodium citrate, adjusted to pH=7, having a total volume of 1 liter of water); 1 X FPB (10 X FPB stock solution = 0.2% Ficoll, 0.2% polyvinylpyrrolidone, and 0.2% bovine  
10 serum albumin); 50% Formamide; and, 100 µg/ml denatured salmon sperm DNA. Pre-hybridization was carried out for 20 hours at 37°C.

B. Preparation of the Human Lymphoma-associated Virus cDNA  
15 from cell line 10C9

10C9 virus particles were purified from 10C9 cell culture supernatants which were clarified of cellular debris by centrifugation (10,000 X g, 15 minutes at 4°C), concentrated 100-fold by Amicon™ ultrafiltration, and pelleted through a  
20 25% sucrose cushion/PBS (100,000 X g, 4°C for 90 minutes). The resultant pellet was resuspended in Dulbecco's Phosphate Buffered Saline (PBS), homogenized, and overlaid on a 15-60% sucrose density gradient, centrifuged for 16 hours at 100,000 X g, 4°C; the virus band at 1.15-1.18 g/ml was collected,  
25 diluted with PBS and pelleted by ultracentrifugation to remove the sucrose (100,000 X g, 4°C for 90 minutes).

The viral RNA was extracted from the particles by the method of Chomzynski et al. and was then used as template in random-primed cDNA synthesis (kit obtained from Boehringer-  
30 Mannheim). The cDNA molecules were blunt-ended using the Klenow fragment of DNA polymerase I and all four deoxyribonucleotides (Maniatis et al., page 113). The A linkers (see Figure 2-A) were then ligated to the ends of the cDNA mole-

cules using T<sub>4</sub> DNA ligase (New England Biolabs). These cDNA molecules were then digested with NruI to eliminate any linker dimers which had been formed in the ligation reaction.

The 10C9 cDNA library was then amplified using the sequence-independent single primer amplification method described in Examples 2 and 6.

C. Hybridization of the 10C9 cDNA library to the BLV genome

The pre-hybridized filter described above was then hybridized for 36 hours at 37°C with the 10C9 cDNA library under the following conditions: 50 mM Tris HCl, pH=7.5; 1 mM EDTA; 0.5% SDS; 3 X SSC (20 X SSC stock solution = 175.3 g NaCl, 88.2 g sodium citrate, adjusted to pH=7, having a total volume of 1 liter of water); 1 X FPB (10 X FPB stock solution = 0.2% Ficoll, 0.2% polyvinylpyrrolidone, and 0.2% bovine serum albumin); 50% Formamide; 10% Dextran Sulfate; and, 100 µg/ml of the 10C9 amplified cDNA library.

After the hybridization period the filter was transferred to a clean 1.5 ml microtube and the following series of washes performed:

1. 100 µl of 2 X SSC and 0.5% SDS, at room temperature for fifteen minutes.
2. Repeat step 1.
3. 100 µl 2 X SSC and 0.1% SDS, at 60°C for 30 minutes.
4. Repeat step 3.
5. 100 µl 0.5 X SSC and 0.1% SDS at 60°C for 30 minutes.
6. Repeat step 5.
7. Elution of the bound cDNAs was achieved by washing with 100 µl of 0.1 X SSC and 0.1% SDS at 95-100°C for 15 minutes.
8. Repeat step 7.

The eluents were run on a conventional agarose gel, transferred to nitrocellulose and hybridized with an IL-5 specific probe (Campbell et al; Azume et al); elution of IL-5 specific sequences was seen only in the eluent of steps 7 and 8.

The eluents were amplified as described in Example 7C except using the A primer.

#### Example 10

#### 10 Characterization of 10C9 Virus cDNAs homologous to the BLV genome

The amplified selected 10C9 cDNAs present in the eluents of Example 9 were then cloned to create two libraries: one in lambda gt10 (Protophone™ lambda gt10 system, Promega) and one in lambda gt11 (Protophone™ lambda gt11 system, Promega). The linkers used in the construction of the amplified cDNAs also contained an EcoRI site (see Figure 2) which allowed for direct insertion of the amplified cDNAs into both lambda gt10 and gt11 vectors.

20 The fusion proteins expressed by the lambda gt11 clones were screened with antibody probes (Young et al.). Screening with bovine anti-BLV ( $\alpha$ -BLV) detected hundreds of positive clones from  $4 \times 10^4$  plaques screened. Screening with  $\alpha$ -p24 detected 15 positive plaques out of  $10^4$  screened, these positives were plaque purified and cross-screened with  $\alpha$ -BLV, 4/15 of these clones were positive with both anti-sera. Screening with  $\alpha$ -gp62 detected 4 positive clones which were all also positive when cross-screened with  $\alpha$ -BLV.

30 The lambda gt10 cDNA library was screened with a full length molecular clone of BLV. Plaque hybridization screening (Woo) of lambda gt10 clones containing approximately 600 base

pair inserts yielded 76 positives out of  $2 \times 10^4$  plaques screened.

While preferred embodiments, uses, and methods of practicing the present invention have been described in detail, it  
5 will be appreciated that various other uses, and methods of practice as indicated herein are within the contemplation of the present invention.

## WHAT IS CLAIMED IS:

1. A method of cloning cDNA species which are homologous to a region of contiguous genomic DNA and are selected from a mixture of cDNA species, said method comprising,
  - (a) obtaining the region of interest of contiguous genomic DNA,
  - (b) preparing cDNA species to contain end-terminal priming sequences,
  - 10 (c) isolating single-stranded cDNA species on the basis of their hybridization to the region of contiguous genomic DNA,
  - (d) mixing the isolated single-stranded cDNA species with DNA polymerase, all four deoxyribonucleotide triphosphates, and primers homologous to the cDNA end-terminal priming sequences,
  - 15 (e) reacting the mixture under conditions to produce sequence-independent amplification of the single-stranded cDNA species, and
  - 20 (f) cloning the amplified cDNA species into a vector.
2. The method of claim 1, wherein the region of contiguous DNA contains a region of interest, and said obtaining includes (a) size fractionating a DNA section containing the region of interest, (b) identifying the fraction containing the region, (c) isolating the DNA from the identified fraction, and (d) digesting the isolated DNA with a restriction endonuclease, to cut the region of interest into a plurality of smaller fragments.



3. The method of claim 2, wherein the DNA containing the region of interest is identified in the fraction by use of primer-initiated amplification in the presence of primers which are homologous region.

5

4. The method of claim 1, wherein the contiguous genomic DNA section is bound to a solid support.

5. The method of claim 1, wherein the contiguous genomic DNA section is contained in yeast artificial chromosomes.

6. The method of claim 1, wherein the contiguous genomic DNA section represents the bovine leukosis virus genome and the mixture of cDNA species is a cDNA library made from mRNA obtained from a cell line infected with a virus derived from cell line 10C9.

7. The method of claim 1, for use in identifying cDNAs which correspond to genes located in the same chromosomal region as a known gene, wherein fragments containing said contiguous genomic DNA region are obtained by,

(a) preparative size fractionating of genomic DNA fragments containing the known gene,

(b) ligating linkers to the genomic DNA fragments, containing the known gene, which are useful as primers for sequence-independent amplification,

(c) mixing the DNA fragments with DNA polymerase, all four deoxyribonucleotides, and primers homologous to the linkers present on the ends of the DNA fragments, and

(d) reacting the mixture under conditions to produce sequence-independent amplification of the DNA fragments.

8. The method of claim 7, wherein the preparative size fractionating of genomic DNA fragments containing the known gene involves identification of the DNA fragments of interest by adding primers homologous to the known gene, DNA polymerase, and all four deoxyribonucleotides, to the gel matrix and treating the gel matrix under conditions which promote amplification of the region of the known gene defined by the primers.
9. The method of claim 1, wherein the known gene is a growth factor or growth factor receptor gene.
10. The method of claim 9, wherein the known gene is an interleukin.
11. The method of claim 10, wherein the known gene is IL-5 and the mixture of cDNA species is a T-cell cDNA library.
12. The method of claim 9, wherein the known gene is erythropoietin and the mixture of cDNA species is a renal cDNA library.
13. The method of claim 1, wherein said preparing of cDNA species to contain end-terminal priming sequences further includes the amplification of the cDNA species by mixing the single-stranded cDNA species with DNA polymerase, all four deoxyribonucleotide triphosphates, and primers homologous to the cDNA end-terminal priming sequences, and reacting the mixture under conditions to produce sequence-independent amplification of the single-stranded cDNA species.

14. The method of claim 1, wherein said preparing of cDNA species to contain end-terminal priming sequences includes the cloning of the cDNA species into a suitable vector and using the known 5' and 3' vector sequences, which flank the cDNA insert, as end-terminal priming sequences.

15. The method of claim 14, wherein said preparing further includes the amplification of the cDNA species by mixing the single-stranded cDNA species with DNA polymerase, all four deoxyribonucleotide triphosphates, and primers homologous to the cDNA end-terminal priming sequences, and reacting the mixture under conditions to produce sequence-independent amplification of the single-stranded cDNA species.

16. A method of cloning cDNA species which are homologous to a region of contiguous genomic DNA and are selected from a mixture of cDNA species, said method comprising,

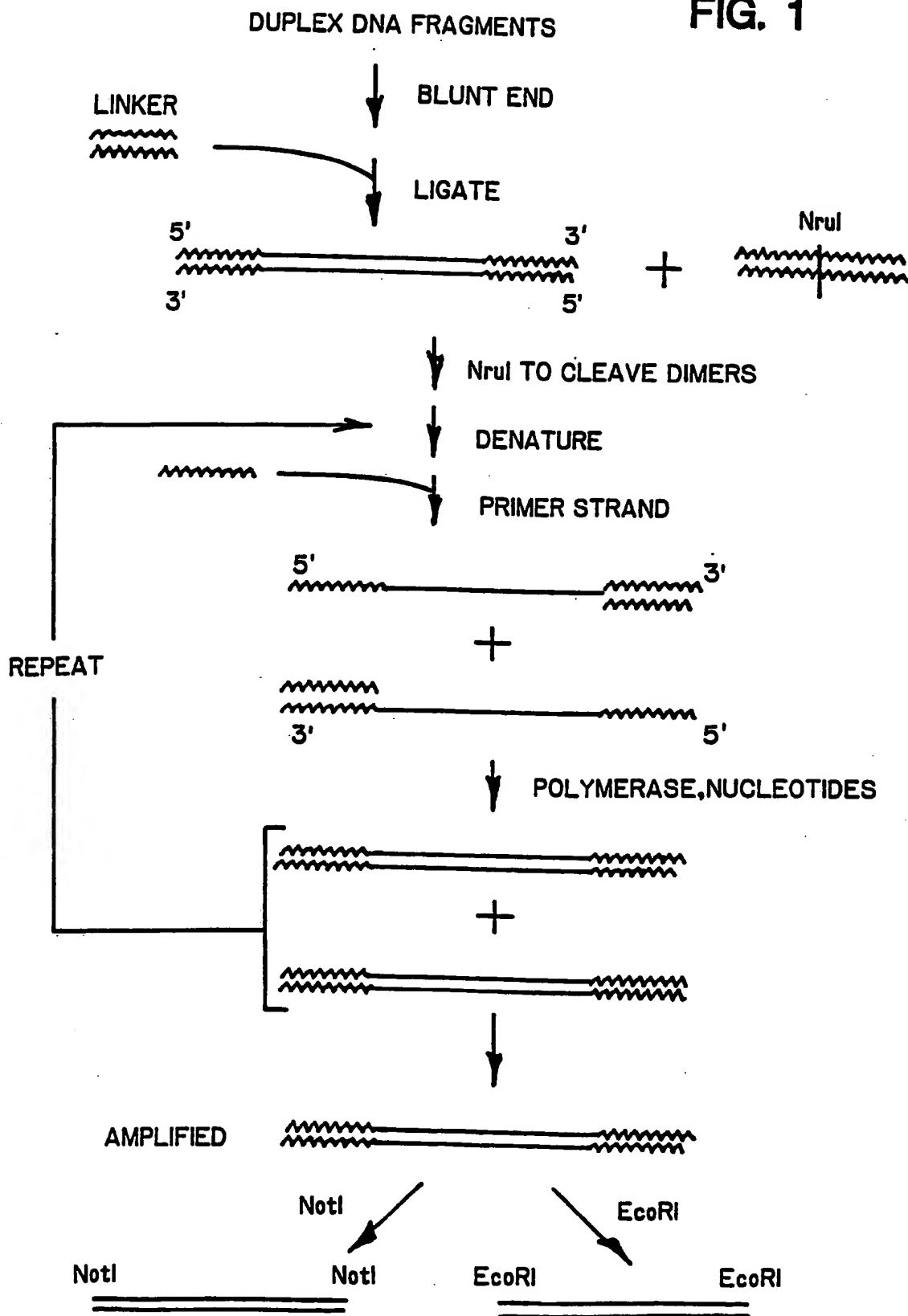
- (a) preparing cDNA species to contain end-terminal priming sequences,
- (b) isolating single-stranded cDNA species on the basis of their hybridization to a first set of selected genomic fragments,
- (c) mixing the isolated single-stranded cDNA species with DNA polymerase, all four deoxyribonucleotide triphosphates, and primers homologous to the cDNA end-terminal priming sequences,
- (d) reacting the mixture under conditions to produce sequence-independent amplification of the single-stranded cDNA species,
- (e) isolating single-stranded cDNA species, from the mixture of step (d), on the basis of their hybridization to a second set of selected genomic fragments,

(f) mixing the isolated single-stranded cDNA species of step (e) with DNA polymerase, all four deoxyribonucleotide triphosphates, and primers homologous to the cDNA end-terminal priming sequences,

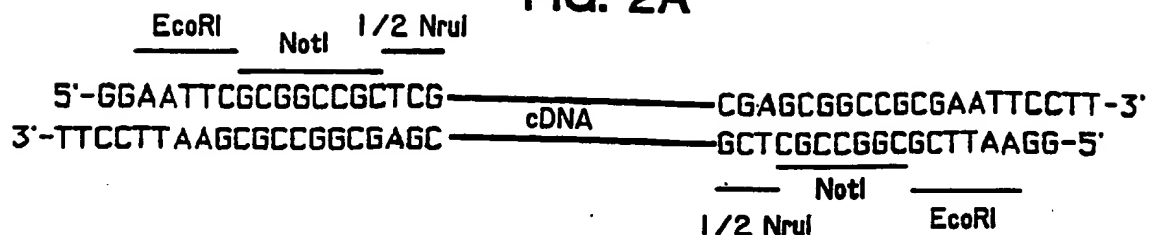
5       (g) reacting the mixture under conditions to produce sequence-independent amplification of the single-stranded cDNA species,

         (h) cloning the amplified cDNA species into a vector.

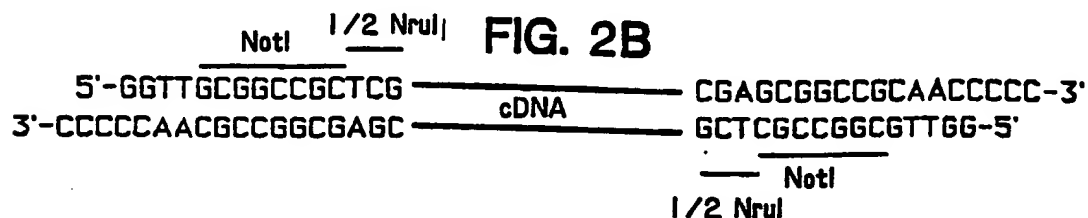
FIG. 1



**FIG. 2A**



**FIG. 2B**



**FIG. 2C**

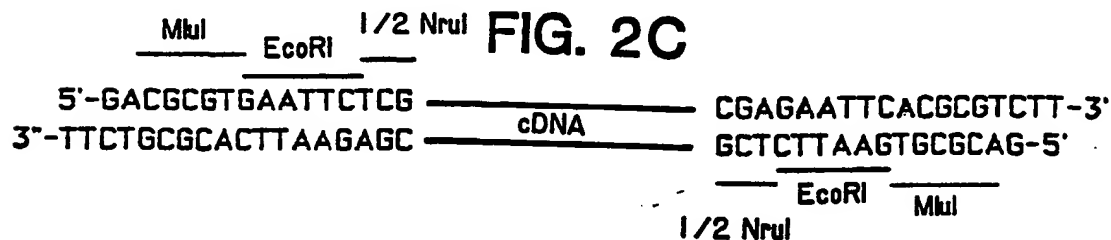
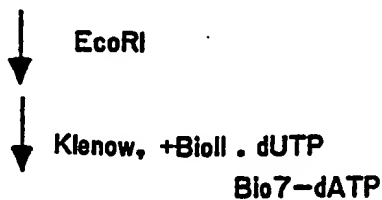
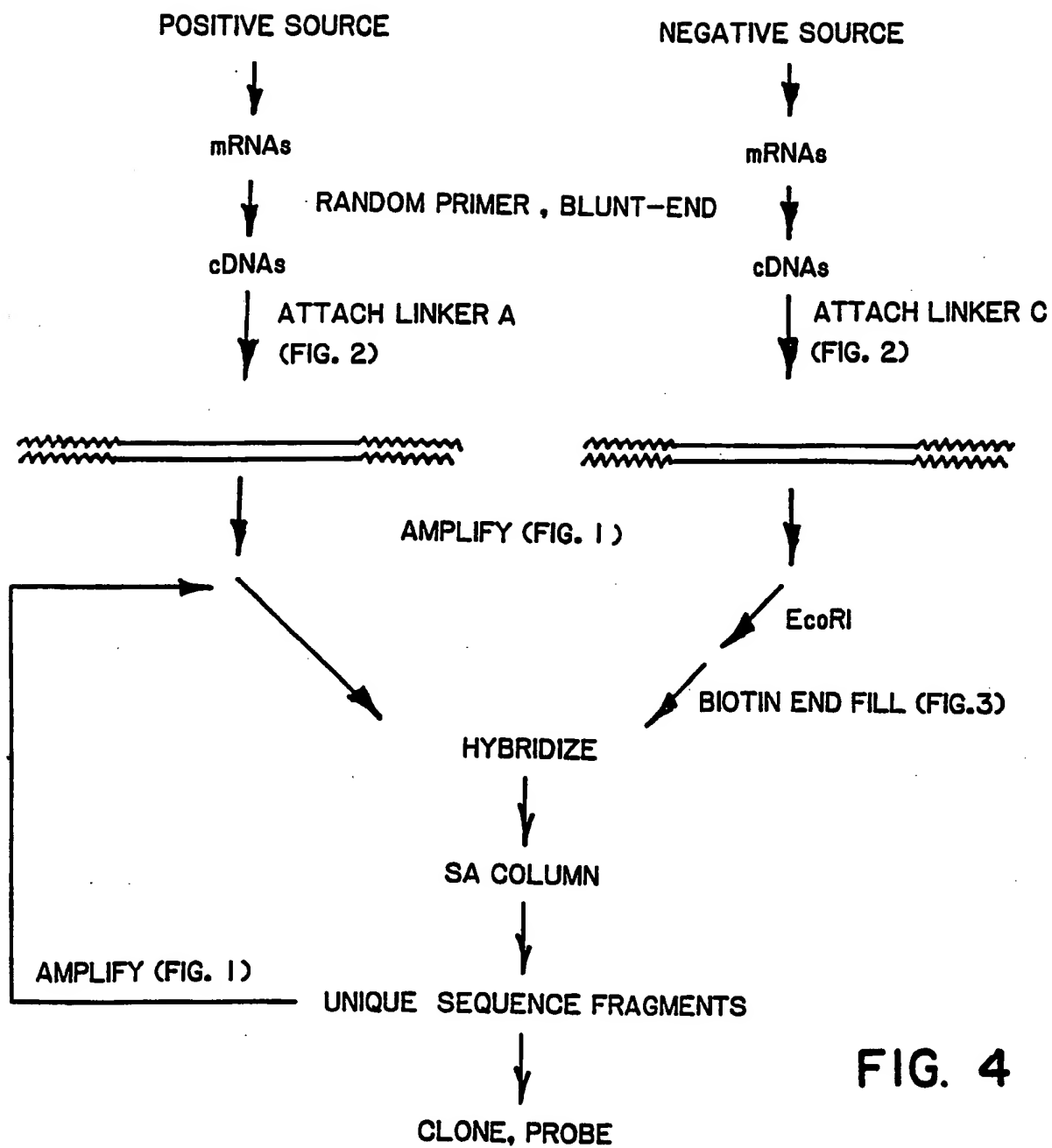


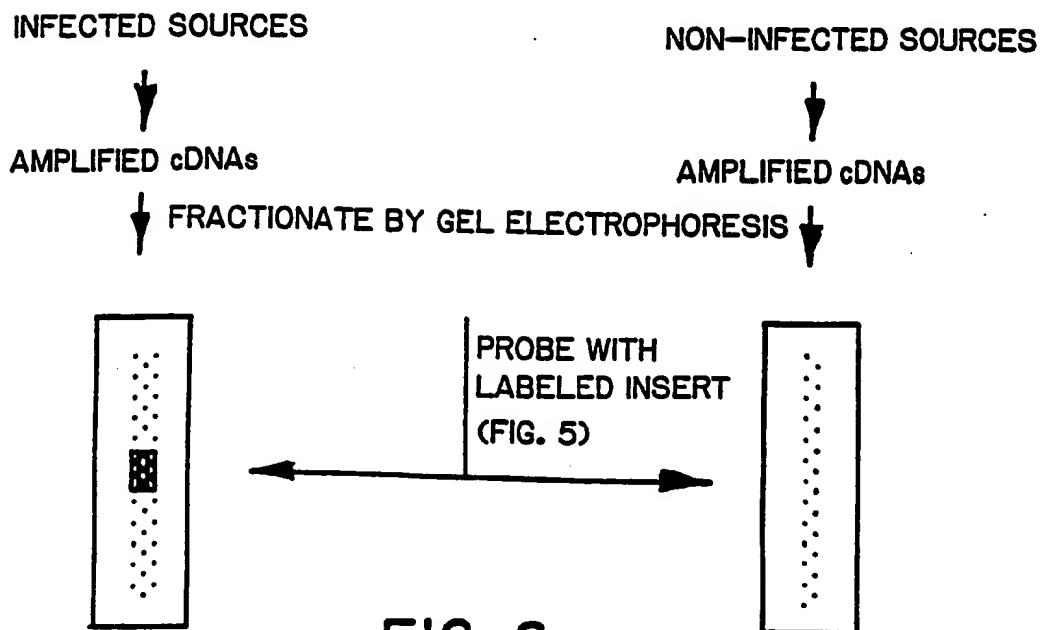
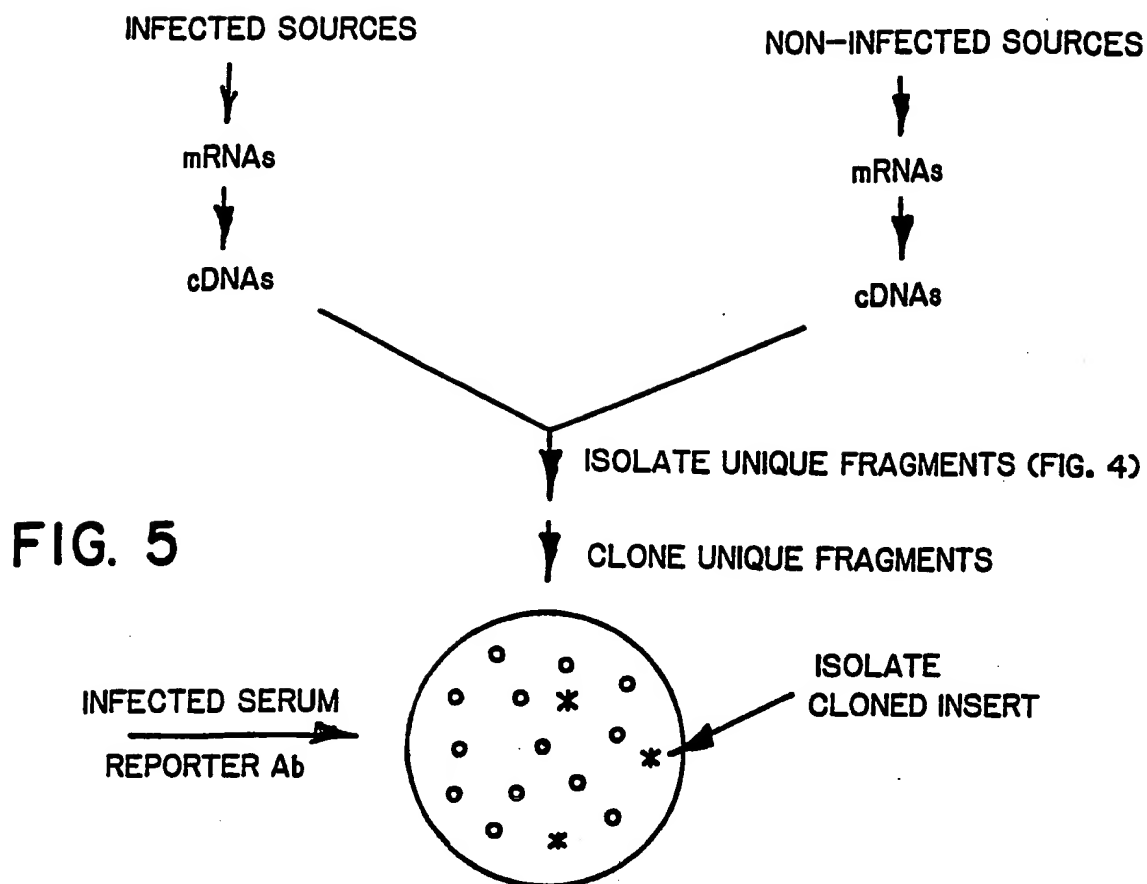
FIG. 3



3/11



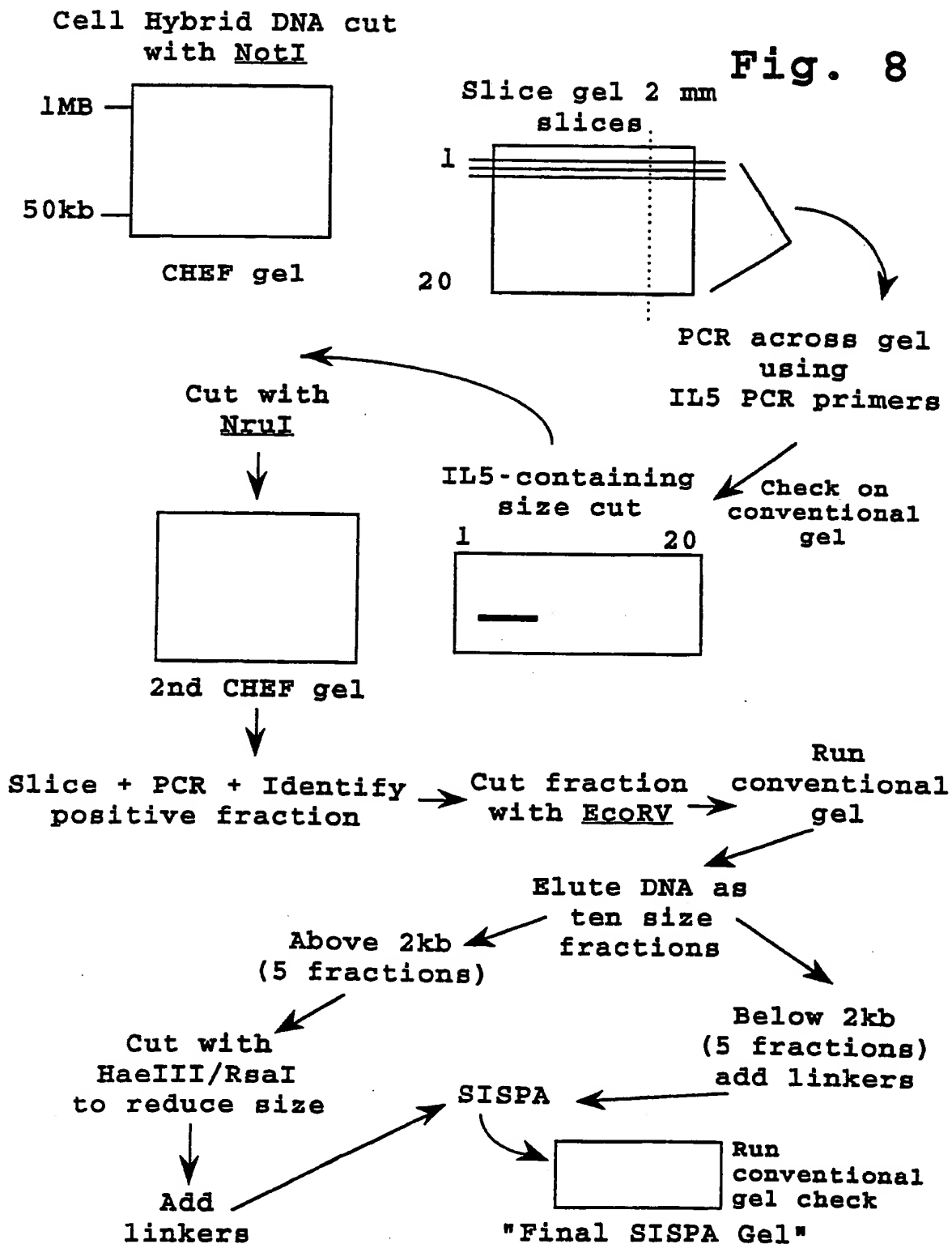
4/11







6/11



Partial Purification using PFGE, PCR and SISPA

7/11



First CHEF Gel (Notl)

**Fig. 9A**



PCR Products

**Fig. 9B**

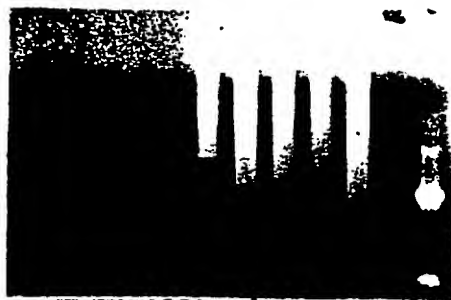


PCR Products Detected by Hybridization

**Fig. 10**

**SUBSTITUTE SHEET**

8/11



Final SISPA Gel

Fig. 11



Eluted SISPA  
cDNAs

Fig. 12

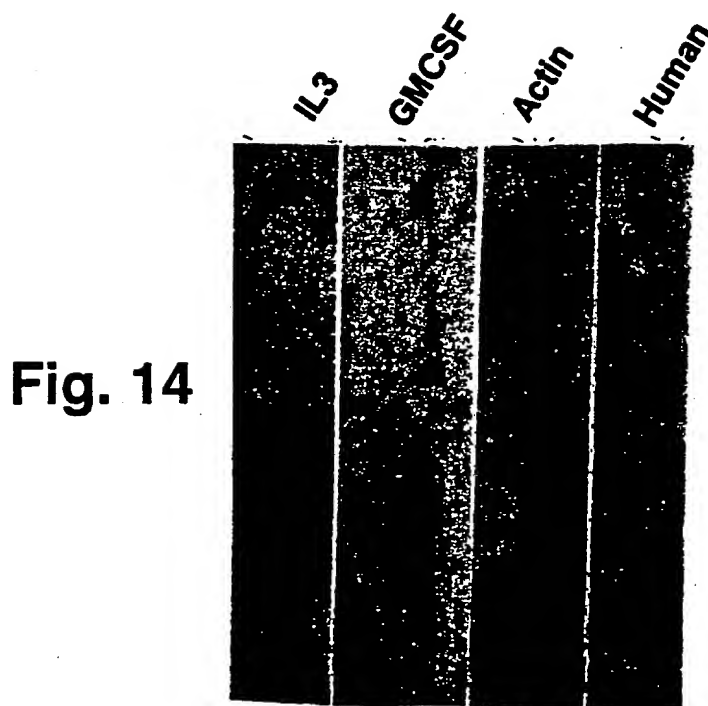
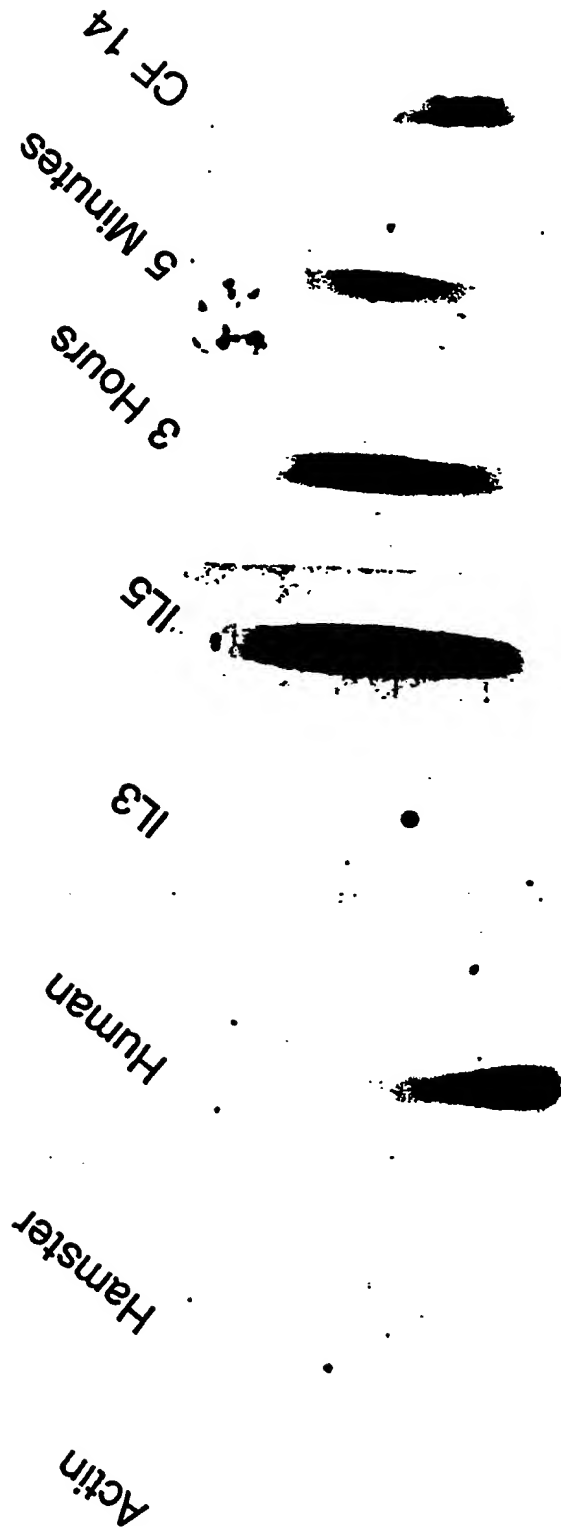


Fig. 14

IL3/GMCSF Eluted cDNAs  
Hybridized with Various Probes

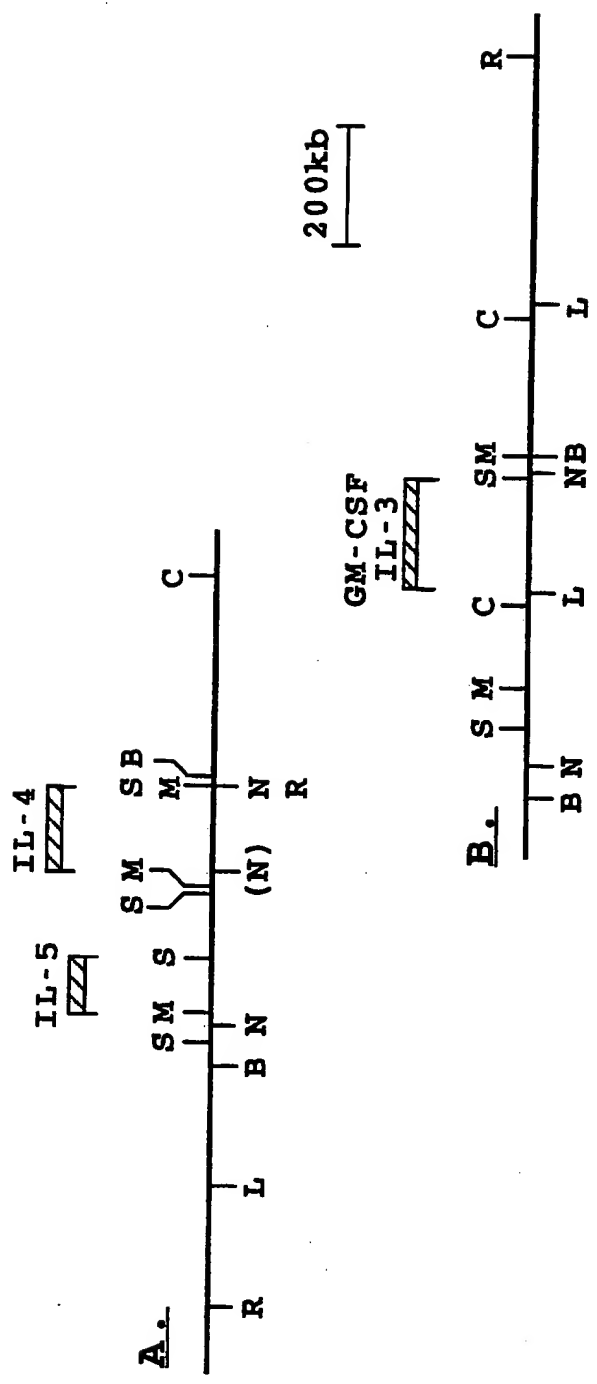
SUBSTITUTE SHEET

9/11



IL5 Eluted cDNAs Hybridized with Various Probes

**Fig. 13**



Long-Range Restriction Map Surrounding IL4 and IL5

Fig. 15

11/11

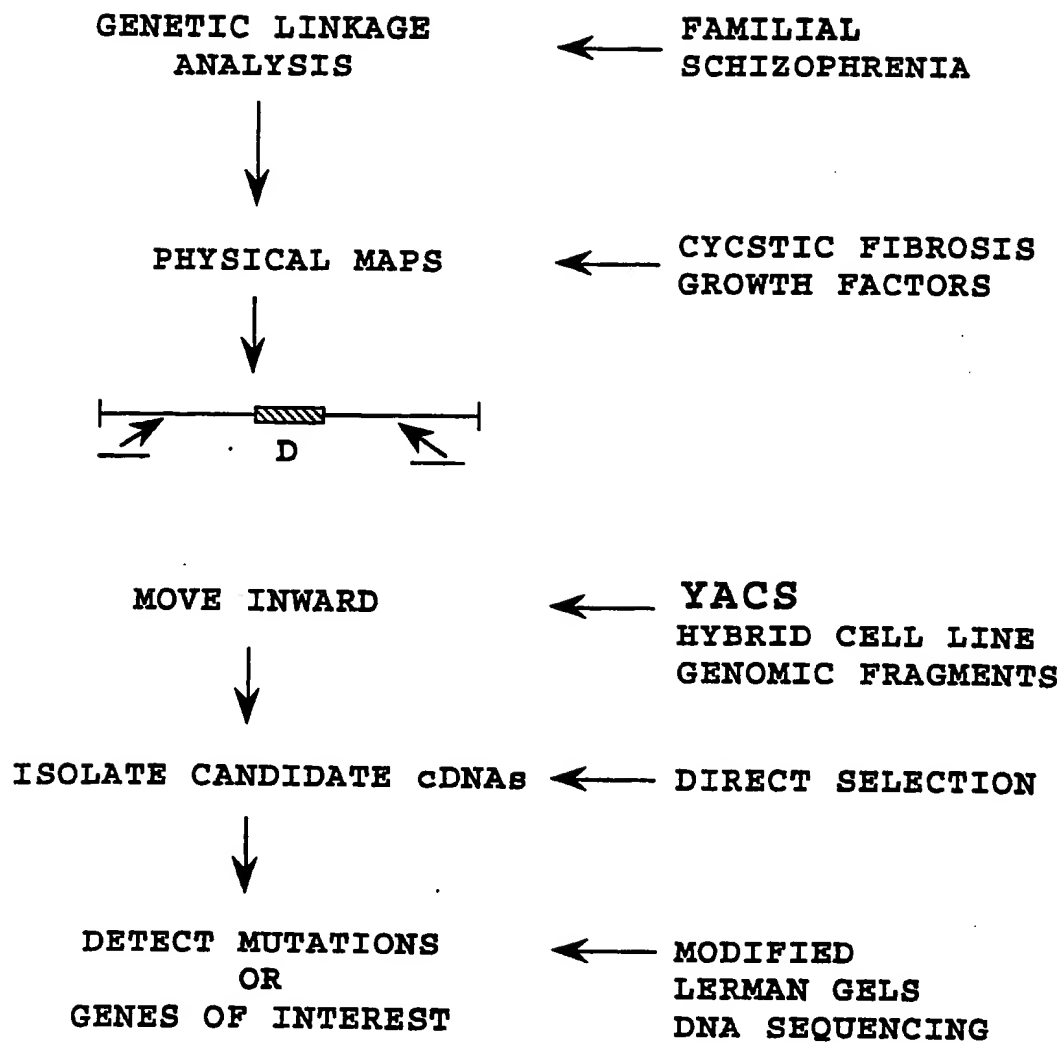



Fig. 16

SUBSTITUTE SHEET

## INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 90/05938

<b>I. CLASSIFICATION OF SUBJECT MATTER</b> (if several classification symbols apply, indicate all) <sup>6</sup>		
According to International Patent Classification (IPC) or to both National Classification and IPC		
Int.Cl. 5      C12N15/10 ;    C12Q1/68		
<b>II. FIELDS SEARCHED</b>		
Minimum Documentation Searched <sup>7</sup>		
Classification System	Classification Symbols	
Int.Cl. 5	C12N ;      C12Q	
Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched <sup>8</sup>		
<b>III. DOCUMENTS CONSIDERED TO BE RELEVANT<sup>9</sup></b>		
Category <sup>10</sup>	Citation of Document, <sup>11</sup> with indication, where appropriate, of the relevant passages <sup>12</sup>	Relevant to Claim No. <sup>13</sup>
O,X	JOURNAL OF CELLULAR BIOCHEMISTRY. Suppl. 0 (13 part E), 3 April 1989, page 293 Abstract WH 160; KIM, J.P. et al.: "Sequence independent single primer amplification (SISPA): utility and application to the detection of enterically transmitted non-A, non-B hepatitis (ET-NANBH)." see the whole abstract ---	1, 13
X,P	WO,A,8912695 (GENELABS INCORPORATED) 28 December 1989 see the whole document (cited in the application) --- -/--	1-16
<sup>10</sup> Special categories of cited documents: <sup>10</sup> "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier document but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such docu- ments, such combination being obvious to a person skilled in the art. "A" document member of the same patent family		
<b>IV. CERTIFICATION</b>		
Date of the Actual Completion of the International Search	Date of Mailing of this International Search Report	
12 FEBRUARY 1991	01 MAR 1991	
International Searching Authority	Signature of Authorized Officer	
EUROPEAN PATENT OFFICE	ANDRES S.M. 	



III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)		
Category °	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim No.
P, X	SCIENCE. vol. 247, 16 March 1990, LANCASTER, PA US pages 1335 - 1339; REYES, G.R. et al.: "Isolation of a cDNA from the virus responsible for enterically transmitted non-A, non-B hepatitis" see page 1336, the legend of figure 2 ---	1-3, 7-8, 13, 16
O, P, X	AMERICAN JOURNAL OF HUMAN GENETICS. Vol. 45 (4th suppl.), 11 November 1989, page A187, Abstract (0732); FUJIMOTO, E. et al.: "The construction of low sequence complexity libraries from partially purified megabase sized DNA." see especially the last paragraph of the abstract ---	1-3, 7-11, 13, 16

**ANNEX TO THE INTERNATIONAL SEARCH REPORT  
ON INTERNATIONAL PATENT APPLICATION NO.**

PCT/US 90/05938  
SA 41767

This annex lists the patent family members relating to the patent documents cited in the above-mentioned international search report.  
The members are as contained in the European Patent Office EDP file on  
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

12/02/91

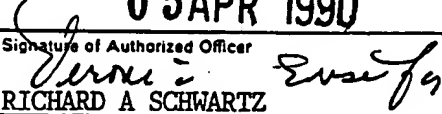
Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO-A-8912695	28-12-89	AU-A- 3852389	12-01-90

EPO FORM P009

For more details about this annex : see Official Journal of the European Patent Office, No. 12/82

# INTERNATIONAL SEARCH REPORT

International Application No. PCT/US90/00434

<b>I. CLASSIFICATION OF SUBJECT MATTER</b> (if several classification symbols apply, indicate all) *		
According to International Patent Classification (IPC) or to both National Classification and IPC IPC(5): C12N 15/09-11, C12Q 1/68, C12P 19/34 U.S. CL: 435/6; 435/172.3; 435/91		
<b>II. FIELDS SEARCHED</b>		
Minimum Documentation Searched ?		
Classification System	Classification Symbols	
US Cl.	435/6, 435/172.3, 435/91	
Documentation Searched other than Minimum Documentation to the extent that such Documents are Included in the Fields Searched *		
Dialog		
<b>III. DOCUMENTS CONSIDERED TO BE RELEVANT *</b>		
Category *	Citation of Document, ** with indication, where appropriate, of the relevant passages †	Relevant to Claim No. ‡
Y	Cell, volume 36, 1984 March, ROHME et al, "Molecular Clones of the Mouse T Complex Derived from Microdissected Metaphase Chromosomes," see pages 783-788, especially results.	1-13
Y	US, A, 4,683,202, MULLIS et al, 28 July 1987 See abstract	1-13
Y	Methods in Enzymology, volume 152, 1987, WU et al, "Adaptors, linkers, and Methylation," see pages 343-345, especially page 345	1-13
Y	Proceedings of the National Academy of Sciences (USA), 1988 December, FROHMAN et al, "Rapid Production of Full-length CDNA'S from Rare Transcripts: Amplification Using a Single Gene-Specific Oligonucleotide Primer," pages 8998-9002, see figure 1, volume 152	1-13
Y	Nucleic Acids Research, volume 9, no. 9, 1981, crampton et al, "The Occurrence of Repetitive Sequences in a Library of Cloned CDNA from Human Lymphocytes," pages 3821-3834, see abstract	11-13
* Special categories of cited documents: † "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier document but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art. "Δ" document member of the same patent family		
<b>IV. CERTIFICATION</b>		
Date of the Actual Completion of the International Search		Date of Mailing of this International Search Report
07 MARCH 1990		05 APR 1990
International Searching Authority		Signature of Authorized Officer
ISA/US		 RICHARD A. SCHWARTZ